АСТРОНОМИЧЕСКИЙ
ЖУРНАЛ

SOVIET ASTRONOMY

AJ

VOLUME 1 NUMBER 1

A Translation
of the
Astronomical Journal
of the
Academy of Sciences of the USSR

(RUSSIAN ORIGINAL VOL. 34, NO. 1)

Published by the
AMERICAN INSTITUTE OF PHYSICS
INCORPORATED

SOVIET ASTRONOMY

AJ

A translation of the "Astronomical Journal" of the Academy of Sciences of the USSR

A publication of the
AMERICAN INSTITUTE OF PHYSICS

Governing Board

FREDERICK SEITZ, *Chairman*

Members

WALTER S. BAIRD
JESSE W. BEAMS
RAYMOND T. BIRGE
R. H. BOLT
J. W. BUCHTA
J. H. DILLON
VERNET E. EATON
HERBERT A. ERF
IRVINE C. GARDNER
S. A. GOUDSMIT
C. KITTEL
WINSTON E. KOCK
R. BRUCE LINDSAY
WALTER C. MICHELS
G. E. PAKE
R. RONALD PALMER
JOSEPH B. PLATT
RALPH A. SAWYER
H. D. SMYTH
JOHN STRONG
GEORGE E. UHLENBECK

Officers and Staff

ELMER HUTCHISSON
Director
HENRY A. BARTON
Associate Director
WALLACE WATERFALL
Secretary and Treasurer
EDWARD P. TOBER
Mgr., Production and Distribution
THEODORE VORBURGER
Advertising Manager
RUTH F. BRYANS
Publication Manager
KATHRYN SETZE
Assistant Treasurer
ALICE MASTROPIETRO
Circulation Manager
EUGENE H. KONE
Director of Public Relations

American Institute of Physics Advisory Board of Russian Translations

ROBERT T. BEYER, *Chairman*

FREEMAN DYSON, DWIGHT GRAY, MORTON HAMERMESH,
DAVID HARKER, VLADIMIR ROJANSKY, HARROLD F. WEAVER

Advisory Editor, Soviet Astronomy–AJ: HARROLD F. WEAVER,
Leuschner Observatory, University of California, Berkley 4, California

"Soviet Astronomy–AJ" is published by the American Institute of Physics, with the cooperation of the American Astronomical Society, for the purpose of making available in English reports of current Soviet research in astronomy as contained in the "Astronomical Journal" of the Academy of Sciences of the USSR. The translation begins with the 1957 issues.

Transliteration of the names of Russian authors follows the system employed by the Library of Congress.

This translating and publishing project was undertaken by the Institute in the conviction that dissemination of the results of research everywhere in the world is invaluable to the advancement of science. The National Science Foundation of the United States encouraged the project initially and is supporting it in large part by a grant. Translation and printing are being handled by Consultants Bureau, Inc.

The American Institute of Physics and its translators propose to translate faithfully all of the scientific material in the "Astronomical Journal" of the Academy of Sciences of the USSR appearing after January 1, 1957. The views expressed in the translated material, therefore, are intended to be those expressed by the original authors, and not those of the translators nor of the American Institute of Physics.

One volume is published annually. Volume 1 will contain all of the scientific material published in the original Russian Volume 34 for 1957.

Subscription Prices:

Per year (6 issues), starting with Vol. 1, No. 1

General:	*United States and Canada*	$25.00
	Elsewhere	27.00
Libraries of non-profit academic institutions:		
	United States and Canada	10.00
	Elsewhere	12.00
Back numbers, all issues		5.00

Subscriptions should be addressed to the American Institute of Physics, 335 East 45th Street, New York 17, New York.

Copyright by American Institute of Physics 1958

SOVIET ASTRONOMY

AJ

VOLUME 1, NUMBER 1 JANUARY-FEBRUARY, 1957

ANNOUNCEMENT

The present issue marks the beginning of the translation of the Soviet "Astronomical Journal." This translation is being published by the American Institute of Physics with the cooperation of the American Astronomical Society, and is supported in large measure by the National Science Foundation.

It further represents the second expansion in the translation program of the American Institute of Physics, which is now carrying out translations of six Soviet journals with a total of approximately 10,000 pages per year. This program, which began in 1955, has been fortunate in the widespread support of the work, not only by the organizations mentioned above, but also by translators, subscribers, and the original editors in the Soviet Union. It is hoped that such cooperation will continue for this new enterprise.

SOVIET ASTRONOMY

AJ

VOLUME 1, NUMBER 1 JANUARY-FEBRUARY, 1957

THE INCLINATION TO THE GALACTIC EQUATOR OF THE GENERAL MAGNETIC FIELD OF THE GALAXY IN THE SOLAR VICINITY

G. A. Shain *

The distribution of the position angles of the planes of preferential vibrations of stellar light is discussed (Figs. 1 and 2). The uniform distribution of θ' at l = = 35-65° and l = 345-355° corresponds to the directions of the spiral arms and leads to a more precise determination of the positions of the latter. Some deviations of θ^1 from 90° are a result of local fields (Figs. 4, 5 and 6).

However, a large part of the deviations is systematic and evidently characterizes the orientation of the general field of the Galaxy in the solar vicinity. In order to determine the latter typical regions with a large number of stars were chosen and the mean values of θ' calculated (Table 1 and Fig. 3). These values are well represented by a sine curve; therefore the plane of the field is inclined to the galactic plane. The angle between the planes is ~ 18°; the planes intersect at $l \sim 15°$ and $l \sim 195°$. The coordinates of the pole, corresponding to the plane of the magnetic field, are $l \approx 105°$, $b \approx 72°$ or $\alpha_0 \approx 12^h$, $\delta_0 \approx +46°$. These coordinates do not coincide with those of the pole of the local system (Gould's belt), nor with those of any other system. The derived plane is not the plane of the magnetic field of the Galaxy as a whole. It evidently characterizes the fluctuations in the direction of the field on a scale of the order of 1000 parsecs (which is much larger than the scale of the local fields), similar to the fluctuations of the spiral arms.

Immediately following the discovery of polarized stellar light it was found that in low latitudes the plane of vibration shows considerable tendency to be parallel to the plane of the galactic equator but that in Scutum (l = 345-355°) and in Cygnus (l = 35-65°) this rule is not followed [1]. It then was seen that these two directions

---

*Deceased.

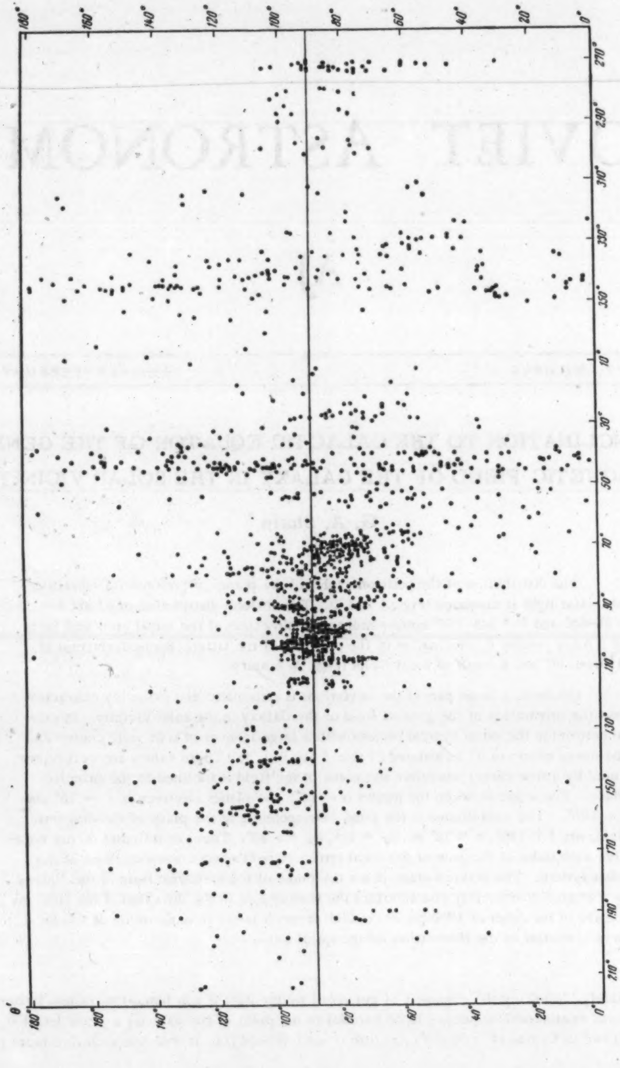

Fig. 1. Distribution of θ' for stars in the zone |b| < ± 6°.

correspond to longitudes where the line of sight is directed approximately along spiral arms (Sagitarrius and Cygnus), and the distribution of angles θ' must be more or less a random one. However, in addition it can sometimes be seen in some of Hiltner's figures [2] that θ' deviates systematically from 90°. An investigation of the slope of the plane of vibration in considerable portions of the sky and of large scale nonuniformities in this slope is of interest for study of the magnetic field structure of the Galaxy, the relationship between diffuse matter and the magnetic field and, finally, the origin of the interstellar magnetic field.

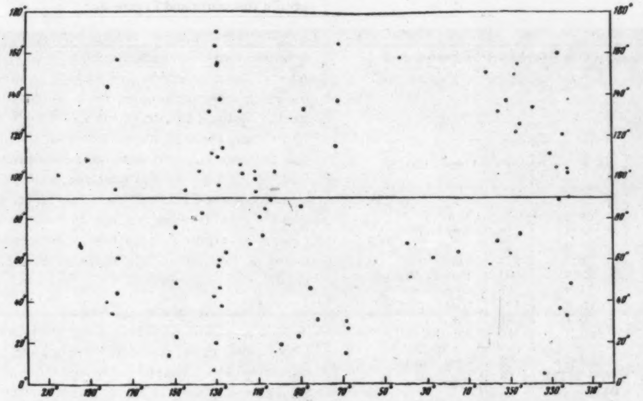

Fig. 2. Distribution of θ' for stars outside of the zone b = ± 8°.

The first question which can arise is whether (within the limits of observational error) the pole which is determined from the magnetic field of the Galaxy coincides with the pole that is determined from the stars. The answer will obviously be of importance in connection with the foregoing problems. For this purpose we shall consider the distribution of position angles θ' which characterize the orientation of the plane of electrical vibration of stellar light with respect to the galactic equator. For study of the longitudinal distribution of polarization it is convenient, as usual, to employ diagrams in which the vectors (P, θ') are plotted. To represent the general picture of distribution of θ' for stars with latitudes ± 8° over 3/4 of the galactic zone we were able to compare only the values of θ' with longitude. In Figure 1 each point represents a star with the coordinates galactic longitude and θ'. For this purpose stars were taken from both of Hiltner's catalogues [3], from Hall and Mikesell's catalogue [4] if not included in [3] and stars of the southern sky from Smith's list in Reference [5] if not included in References [3] and [4]. Stars for which $P \leq 0.3\%$ were not included. In addition, Fig. 2 represents stars of latitudes $b \geq 8°$ with the previously known random distribution of the position angles θ'.

The nonuniform longitudinal distribution of the points in Figure 1 results simply from the real nonuniform longitudinal distribution of hot stars and also possibly from selectivity in the observational data. Of greater interest is the distribution with respect to the ordinate θ' which evidently is entirely independent of the distribution of stars in the sky or of selectivity in the data. Figure 1 clearly reveals two peculiarities.

1) There is more or less uniform distribution of the angles θ' in the interval 0-180° for the longitude intervals 327-347° and 40-50°. In these cases the line of sight appears to be approximately along the spiral arms of the Galaxy. The star density here is considerable and it can be stated that polarization observations definitely reveal spiral arms in Cygnus and Sagittarius. The very marked reduction in the number of stars on both sides of these two directions is another independent argument for the reality of the indicated spiral arms.

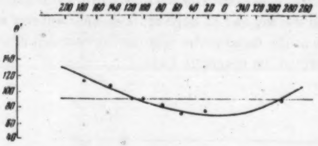

Fig. 3 Average value of θ' for star groups as a function of l

l	θ'	n	Constellation
180—170°	108°2	25	Mon
165—135	104°6	118	Aur-Gem
120—110	91.9	40	
110—90	89.3	274	Pers-Cass
90—66	83.9	303	Pers-Ceph-Cass
58—50	74.3	47	Cyg
50—40	88.2	144	Cyg (spiral)
30—20	76.7	31	Vulp
347—327	89.3	112	Sagittar-Sct (spiral)
301—279	88.0	22	Cent-Norm
274—270	86.6 ± 1°7	18	Crux
	64.8 ± 3.1	12	
300—270	87.4	40	excluding the local field around α Crucis

2) In the majority of regions besides those which correspond to spiral arms in Cygnus and Sagittarius the points are concentrated around $\theta' = 90^\circ$; this is the known tendency of the plane of vibration to be parallel to the galactic equator. However, for the larger part there are systematic shifts to one side or the other. By taking only characteristic regions with a considerable number of stars (for which an average value of θ' is justified) we obtain the table and Figure 3.

Figure 1 and the average values of θ' provide a serious basis for suspecting that the deviations $\theta' - 90^\circ$ can be represented roughly by a sine curve on which are superimposed the individual deviations caused by local smaller-scale fields. We can now confidently speak of the existence of such fields in low latitudes in addition to the general magnetic field of the Galaxy. For example, such local fields can be observed in the "Coal Sack" (Fig. 4) or at $20^h\,5^m + 35^\circ$ (Fig. 5), because in both instances we observe a definite correspondence between the orientation of the field and the structure or size of the nebula.

Although the average values of θ' in Figure 3 depend to a certain extent on the individual groups (represented in the table) any reasonable division into groups results in a picture which is more or less like Figure 3. A glance at Figure 1 (except for sparse stellar regions) shows that Figure 3 correctly represents the data. Comparison with photographs of the sky enables us in some instances to distinguish individual local fields superimposed on the general galactic field. For example, in the region of the Coal Sack we can clearly observe both the general galactic field and the local field which is in the direction, roughly speaking, of the apex of the "bay" to the SE of α Crucis (Fig. 4). For the galactic and local fields the values of θ' are, respectively, 86.6° ± 1.7° (18 stars) and 64.8° ± 3.1° (12 stars).

There is a certain indeterminacy at the other end of the observed region, around 190-210° (Fig. 6). On the one hand, we have here the close group of stars NGC 2323, for which $P \approx 1.2\%$ and $\theta'_{av.} \approx 60^\circ$ (top of Fig. 6). On the other hand, the presence in the photograph of the very elongated emission nebula S172 and of a few dark nebulae elongated in approximately the same direction (at a very large angle ~ 50° to the galactic equator) apparently indicate local fields. This hypothesis is supported by the orientation in the same direction of the plane of light vibration for two stars a little to the west of the emission nebula. It is also possible that in the region of longitudes 190-210° the line of sight is approximately along a spiral arm (opposite to the direction of the arm in Cygnus). When all of these circumstances and especially the small number of stars are taken into account, it is best not to consider this region in connection with the study of the general galactic field.

It can be seen from Figure 4 that the plane of the general galactic field intersects the plane of the galactic equator at 18-20°; the points of intersection (the maximum and minimum of θ') have $l \approx 15^\circ$ and $l \approx 195^\circ$, so that the galactic coordinates of the pole associated with the magnetic field are approximately $l = 105^\circ$ and $b = 72^\circ$, or in equatorial coordinates $\alpha_0 \approx 12^h$ $\delta_0 \approx +46^\circ$. It is of interest that these coordinates do not coincide with the pole of the local system (the Gould belt) of B stars, nor with other systems such as the system of helium stars.

Fig. 4. General field (18 stars) and local field (12 stars) in the Coal Sack.

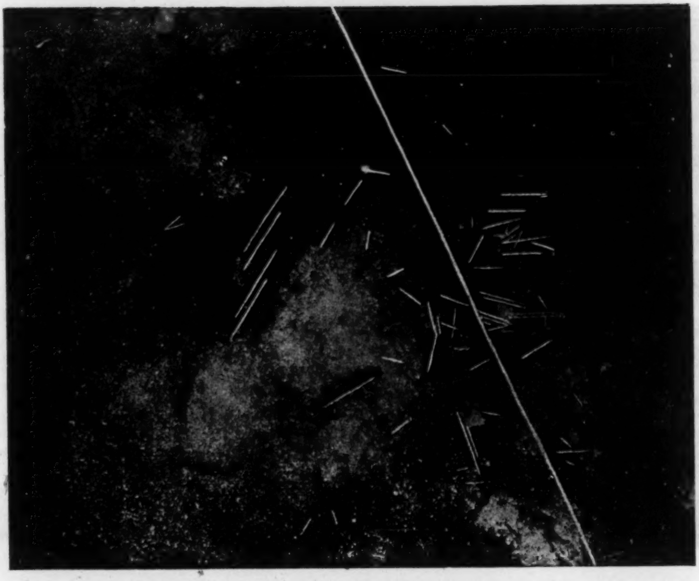

Fig. 5. The region around $20^{h} 5^{m}$ + 35°.

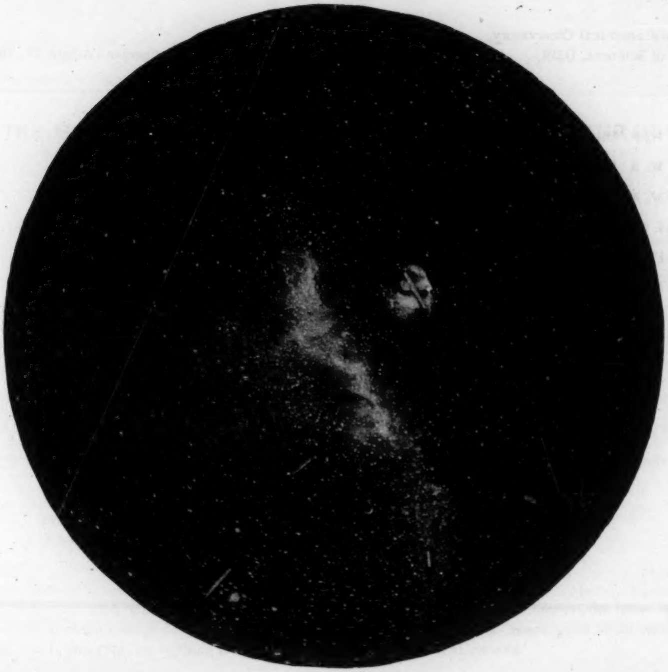

Fig. 6. The region around l = 190-200°.

The coordinates given above for the pole of the magnetic field do not, of course, pertain to the pole of the entire Galaxy. If the general regular field of the Galaxy were to agree fully with the indicated plane the center of the Galaxy would be located a few thousand parsecs from the plane of the field, which is very improbable. If the regular field of the Galaxy is associated with gas clouds it should coincide approximately with the plane of the Galaxy, as shown by 21-centimeter observations. It is most likely that the inclination of the plane of the field shows the existence of fluctuations of the general field, on the scale of about 1000 parsecs, because the separations of most clouds, which through absorption cause the polarization of light from the observed stars do not exceed a few hundred parsecs. The local fields are still smaller and evidently do not exceed a few tens of parsecs.

The investigations of Dutch astronomers have shown that a spiral arm determined by 21 cm radiation does not always lie in the plane of the Galaxy but sometimes forms an angle of as much as 40°. Fluctuations in the structure of the spiral arms is also observed in extra-galactic nebulae. In the present instance we may be dealing with a phenomenon of the same type. It would be interesting to obtain the distribution of gas clouds in the vicinity of the sun for comparison with the plane of the magnetic field which was determined above. The cloud distribution can be obtained by comparing data on interstellar lines, ionized hydrogen regions and 21-cm intensities.

Crimean Astronomical Observatory,
Academy of Sciences, USSR

Received August 27, 1955

LITERATURE CITED

[1] W. A. Hiltner, Astrophys. J. 109, 478 (1949).

[2] W. A. Hiltner, Astrophys. J. 114, 241 (1951).

[3] W. A. Hiltner, Astrophys. J. 120, 454 (1954).

[4] J. S. Hall and A. H. Mikesell, Publ. US Naval Obs. 17, Pt. 1. 1950.

[5] E. V. P. Smith, Sky and Telescope 14, 8, 324 (1955).

[6] G. A. Shain, Astron. J. (USSR) 32, 381 (1955).

THE MORPHOLOGY OF GALAXIES. III. DIFFUSE MATTER IN SPHERICAL STELLAR SYSTEMS

B. A. Vorontsov-Vel'iaminov

Analysis of published data shows that diffuse matter is not at all uncommon in spherical stellar systems consisting of type II populations. It often occurs in the form of dust and, particularly in elliptical galaxies, in gaseous form, and is always present in nuclei of spirals. Observations do not exclude the possibility that diffuse matter may also be present in globular clusters. It is shown that double galaxies similar to M 51, in which the spiral arm of one of the galaxies forms a "bridge" joining it to a smaller, elliptical galaxy, occur fairly often. Elliptical companions of M 51 – NGC 5195 and M 31 – NGC 205 contain diffuse matter in rougly the same proportions as the spiral arms of our own galaxy. Evidently, both of these galaxies are in the initial stages of transformation into spirals. It is probable that spirals evolve from elliptical galaxies as a result of outflow of diffuse matter from their nuclei, if there is a sufficient amount of it. In spherical stellar systems this matter does not appear to be due to accumulation or accretion over a long period of time, but is present from the very start.

1. It is a common belief that spherical stellar systems do not contain diffuse matter, the latter being characteristic of plane systems only. In this connection it is often held that the evolution of plane systems from diffuse matter is probable, but nothing is said about the origin of spherical systems.

This belief is widely held in spite of indications that diffuse matter does occur in a number of spherical systems.

Below, we shall try to summarize and discuss what is known about the presence of diffuse matter in spherical systems such as globular stellar clusters, elliptical galaxies, and nuclei of spirals. Diffuse matter is taken to be present when its density is appreciably higher than that in the surrounding metagalactic space.

2. Globular stellar clusters. There are no indications, so far, of the presence of dust in globular clusters. It would be difficult to detect dust having a density of the order of 10^{-25} g/cc (as in the solar vicinity) in clusters whose dimensions are a few tens of parsecs, but denser clouds, similar to Coal Sacks, could be detected. Thus, although one cannot be completely sure that dust is entirely absent in globular clusters, one can be sure that if it is present it must be distributed just as regularly as the different types of stars in these clusters.

The presence of gas of density of 10^{-24} g/cc, or thereabouts, cannot be detected from spectral absorption lines either, because of the small extent of the clusters.

It would be easier to detect gas by its luminescence, but for this to occur the presence of hot stars is necessary. A star of type O8 has been found in M3 [1], and a star of type B5V in M13.

However hydrogen luminescence can just barely be detected in clusters. It would be brighter if denser accumulations were present in clusters, e. g., gaseous nebulae, which should be close to the single hot stars found in M3. The probability of this is small and hot stars are apparently rare exceptions in clusters. Thus the absence of observations of optical absorption and emission by (ionized) gases in globular clusters cannot be considered as evidence for the absence of gases in them.

If hydrogen were present in clusters, it would be in the neutral state and hence could not be detected from continuous radiospectra which are due to ionized hydrogen. It is true that continuous radio emission is observed over spherical regions of such galaxies as our own and M31. However, phenomenologically, this emission, embracing a region of space comparable with that occupied by the spherical stellar component of our galaxy, does not have the same kind of concentration towards the center as the spherical component itself. Thus the matter which gives rise to the spherical component of radio emission does not have a direct connection with the stars of the spherical component. Consequently, this type of radio emission may be absent in globular clusters because it is not due to stars of the spherical component, but to something else.

If the spherical component of radio emission is due to the bremsstrahlung from relativistic electrons in a magnetic field then in stellar clusters there may not be sufficient sources of such electrons and, in any case, the presence of a sufficiently strong magnetic field is fairly improbable. For this reason (and quite apart from the small dimensions of clusters and the probably small total emission even in most favorable conditions) one would not expect to observe appreciable continuous radio emission from clusters.

The negative result obtained by Mills [2] for globular clusters at a wavelength of 3.5 m must be considered in the light of the above considerations.

There remains the problem of the possibility of detection of neutral hydrogen at 21 cm. I. S. Shklovskii [3] shows that, if the diameter of a globular cluster is 100 parsecs, and the density of gas in it is 10^{-23} g/cc, then at 100° K the radiation from it would be the same as that along the direction towards the galactic center. However, such data are not very useful because it is unlikely that the density of gas, right up to the cluster boundaries, is higher by one order than the density in the spiral arms of the galaxy. Furthermore, the diameters of the largest known clusters are only 60 parsecs, and the stars near their boundaries are very rarefied. If the gas density in the central region of a cluster having a diameter of about 20 parsecs is much less than in the spiral arms of the galaxy, the expected radiation at 21 cm should be weaker by four or more orders of magnitude than that in the direction towards the galactic center and cannot, so far, be detected.

Summing up all the above, we conclude that the existing and possible kinds of observations do not, and cannot, indicate the presence of diffuse matter in globular clusters, but, at the same time, they cannot exclude the possibility that in the dense parts of these clusters there may be gas and dust having densities comparable with those in the galactic plane. It may be assumed that if such matter is present there are no pronounced localizations in it.

3. Elliptical galaxies. Rarefied spherical galaxies of the Sculptor type have not as yet been extensively studied, and we shall not discuss them here. The spherical galaxies known earlier differ from globular clusters in dimensions and mass only, have the same degree of concentration towards the center, and consist of the same stars (including some known bright stars). Elliptical galaxies are redder than globular clusters, and their ratio of mass to luminosity is higher. It is probable that this difference is due to differences in luminosity functions for weak stars. It may, however, be admitted that the increase in the mass — luminosity ratio in elliptical galaxies may be partly due to a greater density of nonluminous diffuse matter. In fact, there are definite indications that diffuse matter is present in elliptical galaxies.

To begin with, according to Mayall [4], the emission line λ 3727 A (OII) was observed in central parts of 20% galaxies of type E. Actually, the percentage of galaxies of type E which contain gas is much higher. It is probable that this spectral region was not sufficiently well explored in all the galaxies which have been studied. According to Page's data [5], among 16 elliptical galaxies which he studied, the line λ 3727 A was observed in five cases, and, in addition, in three of these, bright H_α and intense λ 6584 A (NII) were also noted. The same but weaker lines were observed in galaxies which did not give rise to the line λ3727 A. Thus among the sixteen type E galaxies, the line λ 3727 A was observed in 31%, and the lines H_α and NII alone in a further 31% of these galaxies. Altogether gas emission lines were discovered in 62% of galaxies of type E. Although the data are not very extensive, the large dispersion employed by Page (which weakened the continuous spectrum but made it easier to see the emission lines), and the fact that he covered the red part of the spectrum as well, leads us to suppose that the figure of 62% is much closer to reality than the 20% obtained earlier from observations on the line λ3727 A alone. Thus one may say that gas emission lines are observed in the nuclei of the majority of elliptical galaxies of the usual type.

In our own galaxy the reason for the luminescence of interstellar gas is its irradiation by hot stars, except for rare exceptions. According to Baade [6], about a dozen blue giants are known in elliptical companions of M 31: NGC 205 and NGC 185.

Nothing is known about the presence of hot giants in other elliptical galaxies, since these only appear on special UV photographs. By analogy with globular clusters which only rarely contain hot stars, and the small number of these stars in NGC 205 and NGC 185, it may be supposed that in 38% of type E galaxies without bright lines, the latter are not visible not because the galaxies do not contain gas, but because there is not a sufficient number of hot stars to produce luminescence of the gas. Thus it is probable that gas is present in all elliptical galaxies in appreciable amounts. Gas emission is very considerable in the spherical radio-galaxies NGC 4486 and NGC 5128.

Unfortunately, it is impossible to estimate the amount of gas in type E galaxies since no information is available on the intensity of lines in their spectra, the optical thickness of the gas within them, or whether the gas is luminous or not.

Less is known about the presence of dust in elliptical galaxies. If the dust does not show pronounced localizations, it is very difficult to establish its presence. Dust localizations may even be masked by a thick layer of stars in front of them. Dust clouds come out better on UV photographs, but these have only been taken in the case of the companions of M 31, which is rather convenient since of the four elliptical companions of this gigantic spiral two contain large dust clouds.

Secondly, due to the brightness of the companion of M 31 — NGC 205 of type E5p, on large scale UV photographs [6] of its nucleus, a circular dust nebula can be clearly seen due to its strong absorption, its diameter being 6 parsecs and its distance from the center 100 parsecs on the new distance scale (450,000 parsecs to the group of galaxies M 31). The form of galaxies strongly depends on exposure. Thus in the second photograph of [7] the above dark nebula does not appear in the overexposed middle part of the galaxy. Instead, patches of less dense nebulae on the other (NE) side of the galaxy have merged into a large dark area of 40 x 170 parsecs in size. The above photograph gives the impression that a large number of adjoining clouds, a few parsecs in diameter, surround the nucleus. They are mainly on the periphery of the central part of this galaxy, mainly in the SW part. The major axis of this part we estimate to be 250 to 400 parsecs.

As can be seen from the long exposure photograph (four hours) given in [7], one end of this central ellipse of NGC 205 is cut off along a chord by a dark nebula 40 x 100 parsecs in size. Dark nebulae in NGC 205 are quite visible in the photograph given in the nebular atlas of V. G. Fesenkov and D. A. Rozhkovskii.

The absence of photometric data makes it difficult to make an estimate of the absorption in these clouds or of their mass. Furthermore, lack of knowledge of the exact position of these clouds within the galaxy would still make this estimate difficult even if the photometric data were available.

Assuming that the thickness of the clouds is equal to one half of their mean visible dimensions, and the density is 10^{-24} g/cc, we obtain $10^5 M_\odot$ for their total mass, without taking into account the more rarefied dust in the remaining parts of the galaxy. If, as in the case of the clouds in T Taurus, the amount of gas is 100 times that of dust [8], then the total mass of diffuse matter in NGC 205 is found to be of the order of $10^7 M_\odot$, while $10^5 M_\odot$ is quite possible. Since the mass of this galaxy is probably of the order of $10^8 M_\odot$ we see that the fraction of the total mass taken up by diffuse matter and the stars is similar to that in the spiral arms of our own galaxy.

Baade [6] has discovered about a dozen blue stars of type B and absolute magnitude of 4 (new scale) in UV photographs of NGC 205, and he ascribes the exceptional nature of this fact to the rarity of dust in this area. An examination of this photograph shows that the hot stars are on the periphery of the central region and group themselves near the main dark nebula, a fact which has not thus far been noted.

According to Baade, NGC 205 differs from the usual dense elliptical galaxies by a smaller concentration of stars towards the center, and appears to be a system in transition to a highly rarefied elliptical system of the Sculptor type. On red sensitive photographs it decomposes into stars right into the center. From counts of bright stars Baade [9] has found a major axis of 9' in NGC 205. Reynolds [10], has obtained a major axis equal to 12' on the basis of photometric studies of photographs taken with an exposure of 30^m.

In this connection, the photograph of NGC 205 taken with long exposure and published in [7] is of particular interest.

In this photogrphs the major axis of NGC 205 also has a length of about 12', i. e., about 1500 parsecs. The central ellipse, rapidly weakening towards the edges and visible on other photographs, appears to be only the nucleus of the galaxy. The ratio of its semiaxes is $^1/_2$ and hence NGC 205 was classified as E5. However, in the photograph taken with a long exposure, the nuclear ellipse lies within a weak ellipse in such a way that the major axes do not coincide but form an angle of 30°. The ratio of the axes of the outer ellipse is 1:3.4 which would correspond to E7. However, in a high contrast reproduction made by us of the photograph given in [7] it is clear that the ends of the outer nebula bend in opposite directions, very much like wide spiral arms. It turned out that the author of the photograph (Pease) assumed that NGC 205 is a spiral. Thus, in the classification of spiral galaxies, NGC 205 appears to be in the initial stages of evolution of a spiral galaxy, with the characteristic bright ellipsoidal nucleus relatively larger than in galaxies of type Sa. At the same time, this is the smallest known spiral although on reproductions of photographs taken with a 48" Schmidt objective its length reaches 18', i. e., 2200 parsecs. The structure of NGC 205 is not visible in these photographs. Instead, one gets the impression that it is joined to M31 by a luminous bridge similar to the bridges discovered by Zwicky between some multiple galaxies. The bridge is directed towards the nucleus of M31 and has a width which is somewhat smaller than the minor axis of NGC 205. This impression must be verified with the original negative. It is remarkable that, as was pointed out to us by F. A. Tsitsin, counts of giants carried out by Seifert and Nassau [11] show an offshoot curving outwards, and towards NGC 205, in the outer equal-density curves surrounding this galaxy. This offshoot is much wider than the bridge we expected. The nature of the offshoot and its well defined connection with NGC 205 excludes the possibility that one is dealing here with an accidental bend in the equal density curves of stars belonging to M31. Judging from large scale photographs, spiral arms of NGC 205 do not contain giants, but the inner parts of the spiral arms of M31, and other galaxies of type Sb, do not contain them either.

The lack of coincidence between axes of the visible nucleus and the spiral system of NGC 205 is characteristic of a spiral seen at an angle. The following facts bring NGC 205 close to a spiral: 1) spiral arms, 2) the presence of a nucleus, 3) presence of dust on the nuclear periphery, 4) presence of hot giants although small in number. The presence of these giants within the nucleus is anomalous.

At a distance of 7° from M31 there is a pair of weak elliptical galaxies, NGC 147 and NGC 185, which form a physical system at a visible distance of 58' and also appear to be companions of M31. They both contain a number of globular clusters and resolve into stars. NGC 147 has an extremely small concentration towards the center, and, according to Baade [6], it is free from dust since distant galaxies are visible through it. NGC 185 also has a small concentration towards the center, the latter being intermediate between that in NGC 205 and NGC 147. For this reason they were classified as Ep by Hubble.

In NGC 185, at a distance of 50 parsecs from the center, and still in the dense part, one can see a very dense dust cloud stretched out along the periphery and 8 x 30 parsecs in size. According to Baade [6], there are about a doxen blue giants here, with absolute magnitudes between −3 and −4.

Whether there are any emission lines in the spectra of NGC 185 and NGC 205 we do not know.

After the radio-galaxies NGC 5128 and NGC 1316, and among systems consisting of type II population, the elliptical galaxy NGC 5195, a companion of M51 (Fig. 1), is the richest in dark matter. We note, by the way, that NGC 5195 cannot be an isolated galaxy on which a spiral arm of M51 is accidentally projected as was suggested by Holmberg [12]. It is true that one can see, beyond the edge of the galaxy-companion, a weak chain of hot stars A, which is a continuation of the spiral arm of M51, but this is very short. The spiral arm practically ends where it visibly "sticks into" the companion. The spiral arm of M51 is, at the same time, of the type observed by Zwicky in some multiple galaxies. These "bridges" were shown by Zwicky (from their color only) to consist of hot stars just as in the case of the spiral arm joining M51 and NGC 5195. We have discovered an even more remarkable fact, namely, that such pairs of galaxies are not rare. We found an essentially similar picture for NGC 4485-90. An elliptical companion (also somewhat deformed) is joined by a spiral arm to the main component. Here the main component is of the late Sc type and its equatorial plane is turned towards us much more than M 51. It is possible that NGC 7678 is a similar system judging by the drawing given by Rosse [13]. In the drawing of this spiral galaxy one wide arm ends in a bright circular localization having a diameter equal to the width of the arm. That this is the case is shown in a reproduction of a photograph of NGC 5278-9, found by us later, which has an essentially similar form. NGC 5195, considered by some regular and by others irregular, has the form of an ellipse distorted by two, almost diametrically opposite, regions of large masses of dark matter. In addition, on one side one can see a weak, and reducing to zero, wide continuation of the luminescence of this galaxy. "Analytical photographs" by Zwicky [14] show fairly well that the main mass of stars of NGC 5195 is of the same nature as the nucleus of M 51 itself, i. e., it consists of stars of type II population. Dense areas characteristic of spiral arms and formed by concentrations of giants and gaseous nebulae are not visible in the galaxy-companion.

On the usual "blue" photographs one can see dark matter on the background of the companion which, since it is placed in opposite regions, weakens the radiation emitted by the companion so that the latter appears to be stretched out along the line joining its center to the nucleus of M 51. If the "blue" negative is superimposed on the "yellow" positive of Zwicky [14], the yellowish areas disappear and the dust is even more clearly visible than in the "blue" photograph. Here the companion looks almost like a long rectangle (with torn edges because of the inclusion of dark matter) stretched along the radius of M 51. Both from this picture and the "blue" picture it is quite clear that the end of the spiral arm which leans against the companion consists mainly of dust, with a small number of stars with moderate liminosity, and a very small number of supergiants and their concentrations. The width of this arm is just equal to the diameter of the companion. In the opposite part of the companion one can see a dark patchy nebula a with a protruding area and fibers b, c and d (Fig. 1). The sharp ends of these fibers point away from the nucleus of NGC 5195 and almost at right angles to the spiral arm. For this reason these fibers could hardly be a continuation of the spiral arm. Also, on the opposite side of the companion, dust is more dense than in the spiral arm and hence a part of it may belong to the companion itself and not to the spiral arm as such.

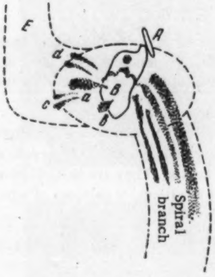

Fig. 1. Schematic drawing of NGC 5195. Dust is shown by the dark patches.

An unusally interesting feature of reproductions of M 51 given in [15] and [16] is a long, thin, and curved canal, 15 parsecs thick (assuming the presently accepted distance of 2.5 million parsecs) which curves round the central dense region of the nucleus of NGC 5195. It passes the center of NGC 5195 at a distance of only 50 parsecs (the diameter of this galaxy is not less than 1500 parsecs). Its length is 250 parsecs and it is perpendicular to the spiral arm. The fact that it curves round the dense patch in the nucleus clearly shows that it belongs to the elliptical galaxy and not the spiral arm of the main component.

On the "yellow" photograph, and the "yellow" negative combined with the "blue" positive, the companion has the form of an ellipse, the major axis of which is perpendicular to the spiral arm and the extension of the companion in " blue" photographs. In the compound photograph one can see, in addition, a wide continuation E of the luminescence of the companion which bends in a way which suggests the beginning of a wide spiral arm. The wedge-shaped fibers cut into the beginning of this would-be spiral arm. At the diametrically opposite end of the compound photograph one can see a similar continuation of luminescence but much shorter and bending in the opposite direction.

The elliptical galaxy rich in dust, NGC 5195, which consists of stars of type II population, and which cannot be called irregular only because its outline is broken up by dark nebulae, contains gas. In its spectrum [5], the emission line H_α can be seen but the lines NII and λ 3727 A are absent, and the lines $H_\beta - H_\delta$ are seen in absorption. It is probable that in this integral spectrum of the galaxy, stellar absorption lines mask the always weaker gas emission lines $H_\beta - H_\delta$. One may expect that these emission lines are due to a small number of hot stars in the central part of NGC 5195. NGC 5195 is evidently similar to NGC 205 in that it contains a considerable amount of dust and shows signs of formation of spiral arms. Both of these galaxies are interesting in that they represent a new type of galaxy which is transitory between elliptical and spiral, and is related to the nuclei of spiral galaxies. The presence of dust in them is very significant. One may assume that if dust is present in an elliptical galaxy (undoubtedly together with cool gas), or when they appear, the formation of spiral arms begins. The degree of development (length and strength of arms) probably depends on the amount of diffuse matter within the nucleus of the galaxy. The amount of diffuse matter in the nucleus may be expected to be proportional to the volume of the nucleus. This does not explain the striking fact that in spite of the difference in absolute dimensions of galaxies, the length of arms in degrees is roughly the same. For example, M31 is eight times as large as M33, but the length of arms in degrees is not very different.

The presence of dust in elliptical galaxies and the nuclei of spirals should be shown in UV photographs (in which absorption is more marked), and with suitable exposure. In the usual photographs these objects in the central parts are usually overexposed, and this masks the dust, which even without this is masked by a thick layer of stars in front.

In the absence of special observations it is difficult to judge the degree of diffusion of dust in elliptical galaxies. Apart from the three type E galaxies described above which contain dust, two of which appear to be revolving spirals, two radio-galaxies of type E with strong bands of dust are known. They are NGC 5128 and NGC 1316. We consider that these radio-galaxies are not colliding galaxies but irregular galaxies of type E, rich in dust and gas, and we shall discuss them in another paper.

4. Nuclei of spiral galaxies. It is known that these do not differ in form from globular clusters and elliptical galaxies, and are comparable to the latter in size. The brighter stars in the systems of all three types are evidently the same both in luminosity and spectrum.

However, nuclei of spiral galaxies differ from globular clusters and elliptical galaxies in having a central plane component rich in gas and dust. Gas and dust concentrate in the spiral arms which in late galaxies is rich in hot giants which cause the gas surrounding them to become luminous and visible.

In addition, in the nearer spirals of later type, supergiants and their concentrations appear in the arms only at some large distance from the nucleus. On the other hand, the dark threads of dust stretch up to the nucleus and often even enter it. For example, according to Baade, in M31 supergiants begin at a distance of 2 kparsecs from the center. According to our estimate, luminous spiral arms begin at a distance of about 550 parsecs from the center, while the sharp ends of the dark fibers and separate dark nebulae approach the center to a distance of 80 parsecs, and many of them are within the nucleus.

Thus concentrations of dust occur, in fact, not only in those regions where there are many giants, but also between stars of smaller luminosity and unknown nature which form the inner and the most dense parts of spiral arms. These concentrations exist also in peripheral areas of nuclei which consist entirely of stars typical of spherical components. Thus elliptical systems which are the nuclei of spiral galaxies contain dust clouds and fibres on their peripheries. Undoubtedly, they also contain gas since the formation of dust fibres can only be thought of in terms of gas currents.

As already pointed out by us in [17], in M33 one finds dust up to a distance of 30 parsecs from the nuclear center. However, supergiants are not found within a distance of 100 parsecs from the center, and its visible nucleus, as we have shown, may not contain (or may contain only a small number) of stars of type II population according to Baade. For this reason, nuclei of spirals similar to M33 will not be considered here. If we included them among the elliptical systems consisting of type II population we would find in them a considerable gas and dust saturation. In all the above ares gas will not be luminous in the absence of hot stars.

The gas-dust currents usually curve in the direction of the spiral arms, are radially placed relative to the nucleus and sharpen on the side of the nucleus, while on the outer side they branch out (and not just widen). Currents and "canals" of dust cannot occur without motion along them, and the question arises what is the direction of motion of matter in them. It is difficult to imagine that gas-dust currents flow into the nucleus from the periphery from all directions. If this were the case, and could somehow continuously occur during an interval of time comparable to the age of the galaxies, the latter would be saturated with gas and dust. This apparently is not the case.

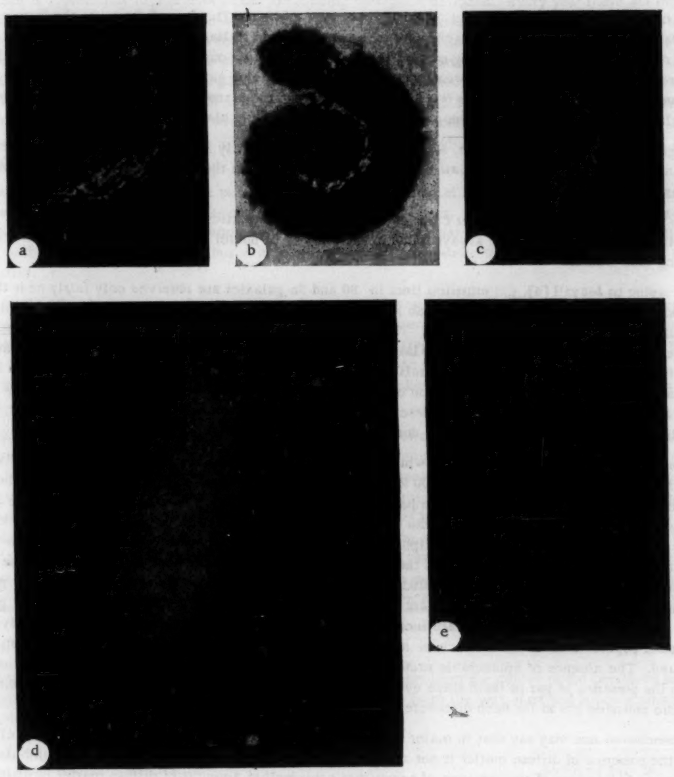

Fig. 2. a) NGC 4485-90; b) NGC 7678; c) NGC 5278-79; d) NGC 205; e) M 51 and NGC 5195.

It is more probable that gas and dust flow out of the nucleus and down the magnetic lines of force in the plane of the system. In that case, however, the nucleus should have a store of diffuse matter which should be reproduced continually within it.

In this connection we wish to point out that there is a spiral of an early type, namely NGC 4826, with arms tightly joined to each other and twisted. An exceptionally strong band (wide and thick) is formed in this spiral and surrounds the periphery of the nucleus giving a strongly asymmetrical form to this large nucleus. Furthermore dark matter cannot be seen in continuous spiral arms.

Thus in some spirals diffuse matter is concentrated in the immediate neighborhood of the nucleus, half of it in its periphery, and not in the spiral arms.

We must also consider the galaxies NGC 2537 and NGC 4038-9. The first of these has the form of a horseshoe consisting of two grains touching each other, i. e., well defined stellar clouds, which in such cases always consist of supergiants. A band consisting of clouds of supergiants comes out of the center of the horse-shoe in the outward direction. All this is on a weak stellar background. In this galaxy, similar in structure, and partly also in form, to spirals of type Sc, there is a bright periphery and the place of a nucleus is taken by weak luminescence. It is possible that the saturation of nucleus with dust is taking place here.

Judging from the drawing of Rosse, NGC 3434 appears to be a fairly regular spiral with very bright inner arms having the form of a ring in the drawing. The central area within the ring, where the nucleus should be situated, is weaker than the outer arms.

One may ask whether in the latter cases one is dealing with a particularly large accumulation of dust in nuclei of spiral galaxies, dust being always found in the vicinity of nuclei of spiral galaxies and sometimes approaching their center.

According to Mayall [4], gas emission lines in S0 and Sa galaxies are observed only fairly near the nucleus, and, evidently, this is also often the case in Sb and Sc galaxies. The percentage of galaxies which emit gas increases in the direction S0 – Sc up to 88%. Thus gas is concentrated in, or in extreme cases luminesces, in the central areas of nuclei just as in elliptical galaxies (when found), while in both cases dust is observed at large distances from the center; in the case of spirals it is usually found on the periphery. Since there is no reason to assume that conditions leading to the emission of the line λ 3727 A (special observations probably show other lines too) arise only in the central parts of these elliptical systems, we may conclude that gas and dust are differently distributed in them: gas in the center, dust nearer the periphery.

A series of spiral galaxies is known in which bright lines in the spectrum of the nucleus are very wide, showing gas motion with velocities up to 4000 km/sec. An inevitable consequence of this should be the outflow of gas from the center to the periphery. Comparing this with the above considerations it appears to be a new and strong argument for the hypothesis that gas can condense into dust, and that for some reason favorable conditions for this happening appear to occur in the periphery, in spite of the fact that one would think that when the central concentrations of gas move outwards towards the periphery their density decreases. The problem of the origin of the luminosity of gas in such systems consisting of type II population still remains unresolved. It is possible that the luminescence of gas is due to hot stars of low luminosity. Nuclei of planetary nebulae in our galaxy belong to this type. Furthermore, there are increasing indications that stars of type 0 of low luminosity are members of type II populations. In the absence of necessary data the amount of gas in the described systems cannot be estimated. The absence of appreciable radio emission from elliptical galaxies does not provide much information on the presence of gas in them since even in the irregular and spiral galaxies which are known to be rich in gas, radio emission has so far been discovered only in single systems.

In conclusion one may say that in major spherical stellar systems, such as nuclei of spirals and elliptical galaxies, the presence of diffuse matter is not at all infrequent. The problem of its presence in globular clusters remains open. Indications of the presence of sometimes considerable amounts of diffuse matter in elliptical galaxies, and of radial motion of dust in NGC 5195 and nuclei of spirals, support the hypothesis that when there is a sufficient store of diffuse matter the formation of spiral arms is due to the outflow of this matter down magnetic lines of force.

5. The origin of diffuse matter in spherical stellar systems. There are three theories of the origin of diffuse matter, or, more precisely, of its origin in the neighborhood of the sun, i. e., near the plane of the galaxy. The earliest theory was that the observed diffuse matter is the remainder of the material from which stars were formed. We showed in 1931 that the loss of gas by stars is sufficient for the formation of stars from diffuse matter thus produced. Recently, some authors have openly, or otherwise, suggested that both stars and the diffuse matter are produced together from some other apparently superdense substance.

However, what can be said about diffuse matter in spherical stellar systems?

By the presence of diffuse matter in a stellar system we mean the case where the density of diffuse matter in the system is appreciably higher than in the surrounding metagalactic space. Every year there are more indications that the density of metagalactic diffuse matter is very nonuniform. However, the passage of galaxies which do not contain diffuse matter (in the above sense) through clouds of metagalactic matter, and the passage of globular clusters through the galactic plane, cannot lead to an enrichment of such spherical systems with diffuse matter since these passages take place at large relative velocities. If such systems contain diffuse matter having a density less (say, by a factor of 1-2) than that in the galactic plane, the take-up of diffuse matter at large relative velocities will evidently not take place either. On the contrary, the denser medium will on collision remove gas from the system which itself will pass freely through the medium and will emerge from it free of gas.

However, one cannot exclude the possibility that in the presence of orbital motion of diffuse masses and spherical systems their encounters will occur at small relative velocities, and with the presence of matter in the system, even with a small excess density, the capture of an additional amount of diffuse matter will be possible. Such cases are probably rare but as a result some spherical systems may occur which are rich in foreign diffuse matter.

A certain amount of gas in spherical systems may be produced from internal resources. Nonstationary objects are present in small numbers in globular clusters. Thus a new star of T Scorpius was observed in M 80 in 1860. A planetary nebula enters into M 5. Ordinary stars will give certain amounts of gas. There is no reason to suppose that the loss of gas by stellar atmospheres is smaller than in the stars of the main sequence. The latter is, however, unknown. According to the hypothesis of V. G. Fesenkov on the evoluatio of stars, this loss is considerable.

In natural conditions, gas will concentrate towards the center of spherical systems, as it would concentrate in the planes of spirals because of their higher density. If one assumes that gas has a density of 10^{-24} g/cc within the central part of a globular cluster having a diameter of 20 parsecs, where stars are still sufficiently closely packed, one obtains a figure of 10^{35} g for the mass of the gas. If the number of stars in a cluster is estimated to be 10^{6} then the mean loss per star is 10^{29} g. If in such clusters stars lose their mass in 10^{11} years (which is smaller by one order than the usual assumed age of these clusters), then the mean mass-loss via corpuscular emission is 10^{18} g per year for each star. Figures between 10^{17} and 10^{20} g per year have been found by various authors for the corpuscular emission from the sun. For this reason corpuscular emission of even stationary stars of average luminosity may, over the assumed length of time, produce with the globular clusters a diffuse medium with a density comparable with that in the neighborhood of the sun.

Similar results are obtained for elliptical galaxies.

In spite of all this, indications of the outflow of diffuse matter from the center to the periphery suggest that this matter is present in the central parts of spherical systems from the very start and is not accumulated in them subsequently.

Shternberg State Astronomical Institute

Received July 26, 1956

LITERATURE CITED

[1] D. Popper, Astrophys. J. 105, 204 (1947).

[2] B. Mills, Austrialian J. Phys. 8, 368 (1955).

[3] I. S. Shklovskii, Astron. J. (USSR) 32, 215 (1955).

[4] W. Baade and Mayall, "Problems of Cosmic Aerodynamics," (Moscow 1953) p. 235.*

[5] T. Page, Astrophys. J. 116, 63 (1952).

[6] W. Baade, Pull. Obs. 10 (Michigan, 1950).

[7] F. Pease, Astrophys. J. 46, 28 (1917).

---

*Russian translation.

[8] A. Lilley, Astrophys. J. 121, 559 (1955).

[9] W. Baade, Astrophys. J. 100, 79 (1944).

[10] Reynolds, Monthly Notices Roy. Astron. Soc. 94, 519 (1933).

[11] C. Seyfert, J. Nassau, Astrophys. J. 101, 179 (1945).

[12] E. Holmberg, Medd. Lund I, 170 (1950).

[13] Rosse, Collected Scientific Papers (London, 1926).

[14] F. Zwicky, Publs. Astron. Soc. Pacific 67, 232 (1955).

[15] G. Abetti, History of Astronomy (London, 1954).

[16] J. Pawsey and Brocewell, Radio astronomy (Oxford, 1955).

[17] B. A. Vorontsov-Vel'iaminov, Astron. J. (USSR) 33, 401 (1955).

THE DETERMINATION OF THE ABSOLUTE MAGNITUDES OF CEPHEIDS AND SUPERGIANTS AND THE ZERO POINT OF THE PERIOD-LUMINOSITY RELATION

Iu. P. Pskovskii

The radial velocities of 182 stars confirm the calibration of supergiants found by Morgan and Keenan from the effect of galactic rotation. The distance modulus for α Per, derived by different methods is discussed and the most probable value is found, $M_V = -4^{m}.0 \pm 0^{m}.3$ (p. e). Using α Per as a mark, the spectral types, absolute magnitudes and luminosity classes of 6 supergiants and 6 subgiants of spectral types F5-G5 and also δ Cep and η Aql were determined from equivalent line widths. A correction to Kukarkin's zero point of the period − luminosity relation was found, $-0^{m}.76 + 0^{m}.40$, from derived median photographic absolute magnitudes. This value is in good agreement with those derived by other investigators using different methods, with corrections, wherever needed, to a unified mean system of stellar astronomical parameters. From all the determinations $-0^{m}.83 \pm 0^{m}.04$ was found as the most probable correction to Kukarkin's zero point.

1. Calibration of the Absolute Magnitudes of Supergiants from the Effect of Galactic Rotation on Radial Velocities

The mean absolute magnitudes of supergiants are determined from the effect of galactic rotation by estimating their mean distance according to Oort's formula

$$V_r = K + Ar\sin 2(l - l_0)\cos^2 b - V_0\cos\lambda, \tag{1}$$

with the usual notation.

This method was applied to supergiants of spectral types B − K by J. Greenstein [1] and R. E. Wilson [2] in 1940; Wilson considered 205 stars. In 1951 Morgan and Keenan [3] used the method to determine the mean absolute magnitudes of supergiants of luminosity class Ia (8 stars) and class Ib (8 stars).

In order to determine the parameters of Oort's formula we used the radial velocities of 308 c-stars of spectral types cB6-cM5 which have been classified by different authors as stars of luminosity classes Ia, Ib and II or as c-stars. The data on radial velocities were taken from Wilson's catalogue [4] and from the list of Feast, Thackeray and Wesselink [5].

We excluded stars with large radial velocities and proper motions and stars in high galactic latitudes.

The stars were divided into groups according to their positions in the sky and the radial velocities and galactic coordinates of each group were averaged. Thirty-one equations of the type of (1) were obtained, the solution os which was obtained by least squares. At the same time the probable errors of the parameters of (1) were obtained.

The results are given in Table 1.

The table shows that our value for V_0 agrees with that obtained by Parenago [6]. Our absolute value for K is greater than that of Wilson because the latter did not drop early B stars with a positive K effect. Our value for the longitude of the galactic center (l_0) is close to the value obtained from infra-red observations (330°).

TABLE 1

Author	Spectral type	No. of super-giants	V_0,km /sec	K,km/sec	Ar, km/ sec	l_0
R. E. Wilson, 1940	cB — cK	205	19.7	−2.0 ± 0.8	16.3	324°4
Our work, 1956	cB6 — cM5	308	14.2 ± 1.4	−3.1 ± 1.0	11.7	331.2
P. P. Parengo, 1951	cO — cM	174	15.7	—	—	—

Wilson [7] and Mel'nikov [8] have shown for a number of objects and Ogorodnikov [9] has shown theoretically that K depends linearly on distance:

$$K = K_0 + K_1^r. \tag{2}$$

We have obtained K_0 and K_1 for 182 stars of luminosity classes Ia, Ib and II in nine distance groups, with distances according to the calibration by Morgan and Keenan [3]. K and Ar for each group were obtained by least squares. Taking A = 19.5 km /sec per kiloparsec and solving a set of 9 equations such as (2) by least squares, we found that K_0 is practically zero (K_0 = 0.5 ± 2.7 km/sec) and K_1 = − 6.5 ± 0.3 (probable error). This value of K_1 is close to that obtained by Mel'nikov [8] for long-period Cepheids if the difference in the assumed value of A is taken into consideration.

After eliminating the K term from (1) we determined the mean weighted distances of 17 star groups distinguished by similarity of spectral type and luminosity class. The following formula was used for the mean weighted distance:

$$r = \frac{\sum (K_1 + \sin 2(l - l_0)\cos^2 b)(V_r - V_0 \cos \lambda)}{A \sum (K_1 + \sin 2(l - l_0)\cos^2 b)^2}.$$

In order to determine the mean absolute magnitudes of the supergiants we used Parenago's method [10] to study light absorption (Greenstein and Wilson used cruder methods) and investigated the dispersion of the distance modulus for stars in each group. The results are given in Table 2.

The probable errors of the mean distances were calculated in the usual way and the probable errors of the mean absolute magnitudes were computed from Equation (30) of Parenago's article on interstellar absorption [10] with certain simplifications.

The last column of the table gives the mean absolute magnitude of the stars in a group when the absolute magnitudes are in accordance with Morgan and Keenan's calibration [3]. The last three lines give the mean absolute magnitudes of Ia, Ib and II luminosity classes with probable error from the mean consistency of values of the mean absolute magnitudes for groups of the table.

This is a confirmation of Morgan and Keenan's calibration of supergiants of luminosity classes Ia and Ib. Stars of class II have a large dispersion of radial velocities and the calibration method cannot be applied to them.

TABLE 2

Spectral types	Mean spectral type	No. of stars	$\bar{m}_V$	$\bar{r}$ kps	$\bar{M}_V$	$\bar{M}_v$ from [3]
B7—A0	B9Ia	10	5^m6	1.27 ± 0.18	$-7^m6 \pm 0.6$	-7^m0
A2—A5	A3Ia	7	5.5	1.54 ± 0.29	−6.6 ± 0.8	−7.2
F0—G0	F5Ia	9	5.0	1.20 ± 0.16	−7.2 ± 0.6	−7.0
G8—M2	M0Iab	8	6.1	0.76 ± 0.36	−5.9 ± 0.5	−6.3
B8—B9	B9Ib	9	6.1	0.66 ± 0.16	−4.7 ± 0.6	−5.4
A0—F0	A4Ib	9	5.3	0.68 ± 0.24	−4.7 ± 0.8	−4.5
F2—F8	F5Ib	8	5.2	0.36 ± 0.15	−4.6 ± 0.9	−4.5
G0	G0Ib	9	5.6	0.40 ± 0.11	−4.0 ± 0.9	−4.5
G2	G2Ib	8	5.8	0.51 ± 0.08	−4.3 ± 0.4	−4.5
G3—K0	G6Ib	13	5.3	0.49 ± 0.20	−4.2 ± 0.9	−4.5
K1—M2	K4Ib	12	4.6	0.40 ± 0.12	−4.4 ± 0.7	−4.5
A5—F6	F2II	14	5.8	0.43 ± 0.18	−3.3 ± 0.9	−2.0
F8—G4	G0II	16	5.6	0.39 ± 0.14	−3.2 ± 0.6	−2.0
G5—G7	G5II	13	6.2	0.38 ± 0.13	−2.7 ± 0.6	−2.0
G8	G8II	11	5.9	0.35 ± 0.16	−2.5 ± 0.6	−2.1
G9—K2	K1II	14	5.9	0.29 ± 0.30	−2.3 ± 0.5	−2.2
K3—M5	K6II	12	5.3	0.28 ± 0.24	−2.7 ± 0.4	−2.4
B7—M2	Ia	32	—	—	−7.0 ± 0.3	−7.0
B8—M2	Ib	70	—	—	−4.4 ± 0.1	−4.5
A5—M5	II	80	—	—	−2.7 ± 0.3	−2.4

No conclusion can be drawn from these results concerning the variation, from one spectral type to another, of the mean absolute magnitude within a luminosity class.

2. Absolute Magnitudes of Five Pseudo-Cepheids and Two Cepheids from Luminosity

The luminosity relation which Hertzsprung discovered in 1905 was thoroughly investigated in 1914 by Adams and Kohlschutter and was used to determine the absolute magnitudes of stars. However, the numerous determinations of absolute spectral magnitudes at the Mount Wilson, Victoria and other observatories from 1923 to 1935 are now obsolete because they did not take interstellar absorption into account.

An important advance was made by Morgan and Keenan's [11] two-dimensional spectral classification and calibration of supergiants with respect to absolute magnitude. At present, however, we have no determination of individual absolute magnitudes of supergiants from luminosity.

The basic difficulty in determining individual magnitudes of supergiants is the almost complete absence of reference stars, that is, supergiants whose absolute magnitudes are known exactly.

Among the pseudo-Cepheids only α Persei belongs to a moving cluster in Perseus and has a cluster parallax (Table 3). However, the determination of the parallax of α Persei from the spectrum-apparent magnitude and color-apparent magnitude diagrams diverges from the group parallax. The discrepancy originates in the inaccuracy of the old calibration of the main sequence [3]. Using the 1956 calibration of the main sequence by Johnson and Hiltner [22], we corrected the distance moduli for α Persei and obtained an average value which is in good agreement with the average obtained from group parallaxes.

The apparent magnitude of α Persei corrected for absorption is $1^m.4$ according to Eggen, Harris and others, so that the absolute magnitude of α Persei is taken as

$$M_v = -4^m.0 \pm 0^m.3.$$

The agreement of the absolute magnitude as determined from the cluster and the calibration of the spectral types F5Ib serves as a proof that α Persei actually belongs to the cluster.

TABLE 3

Method	π	$m_0 - M$	$m_0 - M$ corr.
Group method of [12]	0".008	$5^{m}.4$	
Group method of [13]	0.010	5.0	
Group method of [14]	0.009	5.2	
Group method of [15]	0.007	5.6	

Average for α Persei from group parallax: $5^{m}.4 \pm 0^{m}.1$

Method	π	$m_0 - M$	$m_0 - M$ corr.
Spectrum-magnitude diagram [16]	0".006	$6^{m}.1$	$5^{m}.1$
Spectrum-magnitude diagram [17]	0.0055	6.3	5.5
Spectrum-magnitude diagram [18]	0.004	6.85	5.85
Color-magnitude diagram [19]	0.007	5.85	5.7
Color-magnitude diagram [20]	0.010	5.0	5.0
ν-component of proper motion [17]	0.008	5.4	5.4

Average from distance modulus of cluster $5^{m}.5 \pm 0^{m}.1$.

The trigonometrical parallax of α Persei in Jenkins' catalogue [23] is considerable, being ± 0". 029 ± ± 0".005. But this is an error: the parallax in the catalogue of the Allegheny Observatory [24] is + 0".007 ± ± 0". 005 and not 0." 038 ± 0." 006 as Jenkins assumed in connection with the absolute parallax. Our corrected trigonometrical parallax for α Persei is + 0." 006 ± 0." 004 (mean square error) in Jenkins' catalogue system. Accordingly $M_{trig} = -4^{m}.7 \pm 1.^{m}4$.

For the purpose of determining the absolute magnitudes of five pseudo-Cepheids and two Cepheids we obtained 40 spectrograms of six supergiants, six subgiants and two Cepheids using the 120-cm reflector of the Crimean Astrophysical Observatory of the USSR Academy of Sciences. The spectral dispersion was 35.6 A/mm for H_γ. Ilford Zenith plates were used. An MF-4 microphotometer was used to process this material. The continuous spectrum in the spectrograms was taken from the traces for α Persei in the atlas of Hiltner and Williams [25], for the sun in Minnaert's atlas and for γ Cyg in one of Iwanowska's articles [26].

Table 4 gives the equivalent breadths of 18 lines and of the G band for stars of the program. The asterisks denote lines used in determining the absolute magnitudes and spectral types.

The stellar spectral types and absolute magnitudes were determined from diagrams whose ordinates were logarithms of the ratio of equivalent breadths of pairs of lines and whose abscissas were the absolute magnitudes (visual).

In determining the spectral types it was assumed that the absolute magnitudes of the supergiants are $-4^{m}.5$ (from Morgan and Keenan's calibration), and the absolute magnitudes of the subgiants were calculated from the trigonometrical parallaxes in Jenkins' catalog. The criteria of spectral types were the ratios of equivalent breadths of the hydrogen lines H_δ and H_γ to the lines of ionized metals, as well as the ratio of the equivalent breadths of the lines λ 4077 Sr^+ and λ 4383 Fe.

TABLE 4

Star / λ Å	α Per	θ Cyg	ξ Peg	γ Cyg	υ Peg	β Aqr	ζ Her	ε Aqr
* 4045 Fe	0.86 ± 0.03	0.42 ± 0.10	0.45 ± 0.04	0.90 ± 0.00	0.69 ± 0.04	0.93 ± 0.07	1.06 ± 0.07	0.99 ± 0.10
* 4063 Fe	0.55 ± 0.04	0.23 ± 0.09	0.23 ± 0.03	0.80 ± 0.08	0.38 ± 0.07	0.74 ± 0.21	0.82 ± 0.03	0.97 ± 0.01
* 4072 Fe	0.48 ± 0.04	0.33 ± 0.10	0.29 ± 0.04	0.64 ± 0.16	0.34 ± 0.02	0.57 ± 0.11	1.00 ± 0.01	0.73 ± 0.17
* 4077 Sr$^+$	1.11 ± 0.05	0.30 ± 0.08	0.40 ± 0.06	1.22 ± 0.05	0.54 ± 0.03	1.14 ± 0.21	0.70 ± 0.06	1.13 ± 0.05
* 4102 H_δ	6.81 ± 0.22	3.82 ± 0.15	3.33 ± 0.03	4.05 ± 0.21	2.24 ± 0.03	2.82 ± 0.27	2.90 ± 0.12	2.54 ± 0.03
4171 Ti$^+$	1.35 ± 0.04	0.60 ± 0.04	—	—	—	0.77 ± 0.30	0.84 ± 0.09	—
4215 Sr$^+$	0.86 ± 0.02	0.27 ± 0.06	0.31 ± 0.03	1.06 ± 0.07	0.38 ± 0.13	0.75 ± 0.07	0.50 ± 0.01	0.88 ± 0.20
* 4227 Ca	0.93 ± 0.02	0.40 ± 0.01	0.61 ± 0.10	1.11 ± 0.02	0.69 ± 0.12	1.15 ± 0.07	1.38 ± 0.01	1.37 ± 0.01
4247 Sc$^+$	0.68 ± 0.04	0.26 ± 0.01	0.16 ± 0.02	0.92 ± 0.10	0.20 ± 0.09	0.90 ± 0.09	0.42 ± 0.01	0.92 ± 0.17
4250 Fe	0.50 ± 0.02	0.19 ± 0.01	0.35 ± 0.02	0.66 ± 0.01	0.21 ± 0.10	0.46 ± 0.06	0.52 ± 0.01	0.67 ± 0.07
4272 Fe	0.60 ± 0.03	0.26 ± 0.02	0.34 ± 0.05	0.90 ± 0.06	0.39 ± 0.10	0.77 ± 0.08	0.90 ± 0.03	0.85 ± 0.03
4290 Cr	0.98 ± 0.04	0.31 ± 0.08	0.47 ± 0.16	1.25 ± 0.13	0.70 ± 0.16	1.34 ± 0.06	1.20 ± 0.12	1.54 ± 0.18
* 4325 Fe	0.99 ± 0.03	0.48 ± 0.00	0.41 ± 0.12	0.99 ± 0.07	1.07 ± 0.11	1.30 ± 0.06	1.06 ± 0.01	1.52 ± 0.16
* 4340 H_γ	6.00 ± 0.22	3.19 ± 0.20	2.96 ± 0.05	4.17 ± 0.30	2.47 ± 0.49	3.48 ± 0.34	3.34 ± 0.35	3.15 ± 0.15
* 4383 Fe	1.18 ± 0.10	0.39 ± 0.01	0.49 ± 0.03	1.54 ± 0.08	0.91 ± 0.08	1.82 ± 0.06	1.40 ± 0.10	2.02 ± 0.04
4444 Ti$^+$	1.39 ± 0.06	0.29 ± 0.02	0.35 ± 0.10	2.18 ± 0.23	0.56 ± 0.23	1.50 ± 0.05	0.85 ± 0.24	2.03 ± 0.04
4455 Ca	0.70 ± 0.05	0.35 ± 0.06	0.29 ± 0.03	0.78 ± 0.06	0.61 ± 0.06	0.61 ± 0.02	0.46 ± 0.06	1.01 ± 0.12
4461 Fe	0.77 ± 0.02	0.23 ± 0.01	0.22 ± 0.04	1.07 ± 0.03	0.43 ± 0.05	0.62 ± 0.22	0.46 ± 0.02	0.75 ± 0.02
* G-band	2.36 ± 0.46	1.91 ± 0.30	3.40 ± 0.02	3.49 ± 0.08	4.91 ± 0.08	4.77 ± 0.32	6.25 ± 0.15	7.06 ± 0.04

(continued)

TABLE 4 (continued)

Star / λ Å	β Dra	9 Peg	μ Her	η Her	δ Cep ${}_0P_{50}$	δ Cep ${}_0P_{65}$	δ Cep ${}_0P_{74}$	α Cep ${}_0P_{89}$	η Aql. ${}_0P_{90}$
* 4045 Fe	0.89 ± 0.06	0.87 ± 0.10	0.95 ± 0.01	1.01 ± 0.02	0.97 ± 0.04	0.73 ± 0.11	1.05 ± 0.20	0.64 ± 0.09	0.64 ± 0.09
* 4063 Fe	0.78 ± 0.02	0.60 ± 0.01	0.87 ± 0.03	0.63 ± 0.06	0.93 ± 0.05	0.68 ± 0.13	0.75 ± 0.10	0.59 ± 0.10	0.80 ± 0.12
* 4072 Fe	0.60 ± 0.01	0.77 ± 0.17	0.98 ± 0.10	0.61 ± 0.06	0.68 ± 0.13	0.54 ± 0.15	0.44 ± 0.08	0.33 ± 0.04	0.55 ± 0.13
* 4077 Sr⁺	1.04 ± 0.06	0.98 ± 0.05	0.54 ± 0.00	0.53 ± 0.08	1.02 ± 0.20	0.74 ± 0.19	0.93 ± 0.20	0.77 ± 0.00	1.04 ± 0.15
* 4102 H_δ	2.48 ± 0.17	1.56 ± 0.03	1.69 ± 0.14	1.39 ± 0.03	3.52 ± 0.48	2.18 ± 0.03	3.22 ± 0.30	2.84 ± 0.46	3.55 ± 0.20
4171 Ti⁺	—	0.79 ± 0.16	1.17 ± 0.31	—	—	—	—	—	—
4215 Sr⁺	1.07 ± 0.12	1.14 ± 0.18	0.67 ± 0.02	0.89 ± 0.03	0.68 ± 0.19	0.62 ± 0.02	1.26 ± 0.05	0.73 ± 0.07	0.70 ± 0.04
* 4227 Ca	1.25 ± 0.12	1.58 ± 0.01	1.42 ± 0.10	1.50 ± 0.09	1.40 ± 0.03	1.08 ± 0.06	1.30 ± 0.17	1.19 ± 0.06	1.12 ± 0.11
4247 Sc⁺	0.97 ± 0.17	0.82 ± 0.01	0.40 ± 0.10	0.33 ± 0.06	—	—	—	—	—
4250 Fe	0.57 ± 0.02	0.74 ± 0.04	0.60 ± 0.01	0.50 ± 0.02	—	—	—	—	—
4272 Fe	0.91 ± 0.17	1.40 ± 0.01	0.93 ± 0.15	1.04 ± 0.05	—	—	—	—	—
4290 Cr	1.22 ± 0.18	1.04 ± 0.32	0.59 ± 0.05	0.69 ± 0.07	—	—	—	—	—
* 4325 Fe	1.47 ± 0.22	1.99 ± 0.16	1.82 ± 0.20	1.15 ± 0.05	1.50 ± 0.15	0.97 ± 0.28	1.28 ± 0.17	1.22 ± 0.07	1.15 ± 0.01
* 4340 H_γ	2.92 ± 0.07	1.89 ± 0.05	1.92 ± 0.13	1.60 ± 0.12	2.74 ± 0.12	2.20 ± 0.36	3.26 ± 0.08	4,15 ± 0.37	3.59 ± 0.38
* 4383 Fe	1.72 ± 0.29	2.27 ± 0.04	1.66 ± 0.01	1.80 ± 0.20	1.66 ± 0.07	1.29 ± 0.22	1.53 ± 0.10	1.57 ± 0.09	1.26 ± 0.11
4444 Ti⁺	1.95 ± 0.15	1.86 ± 0.04	0.88 ± 0.09	1.02 ± 0.08	1.75 ± 0.23	1.36 ± 0.30	1.62 ± 0.24	1.61 ± 0.09	1.56 ± 0.48
4455 Ca	1.01 ± 0.09	1.30 ± 0.09	0.99 ± 0.39	0.49 ± 0.03	—	—	—	—	—
4461 Fe	0.78 ± 0.01	1.05 ± 0.02	0.65 ± 0.03	0.49 ± 0.02	—	—	—	—	—
* G-band	6.29 ± 0.36	8.90 ± 0.04	5.95 ± 0.17	6.64 ± 0.77	8.15 ± 0.90	5.35 ± 0.80	6.20 ± 0.40	3,45 ± 0.020	8.42 ± 0.51

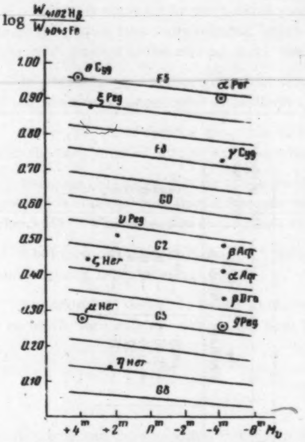

Fig. 1. Grid for the determination of stellar spectral types from the logarithm of the ratio of equivalent breadths of λ 4102 (H_δ) and λ 4045 (Fe) in the spectra of these stars. The symbol ʘ denotes log (W_{4102}/W_{4045}) in the spectra of the reference stars and the dots denote the same in the spectra of other stars.

The reference stars were θ Cyg and α Per (with the spectral type taken as exactly F5) and μ Her and 9 Peg (with the spectral type taken as exactly G5). Thirteen pairs of lines were selected for which the change of the ratio of equivalent breadths of a pair of lines in going from a star of one spectral type to a star of another spectral type is proportional to the change of spectral type from one star to the other. This requirement corresponds to parallelism of the lines which connect points that correspond to ratios of equivalent line breadths in spectra of the same type and to equality of the distances between parallel lines through points representing values of the ratios for the types F5, F6, . . . G5 (Fig. 1). By means of this grid constructed according to equivalent line breadths of the reference stars we determined the spectral types of the stars in Table 5. For determination of the absolute magnitudes the reference supergiant was α Per ($M_V = -4.^{m}0 + 0.^{m}3$). The grid was analogous to that constructed for spectral types. The differences between the logarithms of the ratios of equivalent breadths for line pairs in the spectra of θ Cyg and μ Her provided a basis for ten intervals of the grid; the line connecting points of the drawing corresponding to the logarithm of the ratio of equivalent breadths of line pairs for stars of a single spectral type (F5): θ Cyg and α Per provided the direction parallel to which the grid was plotted. Fifteen pairs of lines were used: the ratios of the equivalent breadths of λ 4077 Sr^+ and λ 4444 Ti^+ to the equivalent breadths for unionized elements.

Table 5 gives our values of the absolute magnitudes with the probable errors from internal consistency. The table shows that α Per, β Aqr and 9 Peg must be assigned to luminosity class Ib-II. Comparison with the spectral types obtained by Gathier[28] confirms our conclusion.

The spectral types of Cepheids were also determined by a different method, which is the relation between the equivalent line breadth and the spectral type of supergiants. This relation was plotted from the data in Tables 4 and 5 for $H_\delta H_\gamma$, λ 4383 Fe, λ 4227 Ca, λ 4325 Fe and the G-band.

From the first three lines the spectral types were earlier than by the other criteria (the Adams-Russell phenomenon) but the spectral type obtained from all criteria is close to the average by the first method.

The median absolute magnitudes of δ Cep and η Aql were obtained from the luminosity curve and Eggen's color index [29]. The average of four values for the median visual absolute magnitude of α Per is $-3.^{m}41 \pm$ $\pm\ 0.^{m}04$ (probable error from internal consistency of the four values) and the median visual absolute magnitude of η Aql (from a single phase) is $-4.^{m}35 \pm 0.^{m}40$.

The median photographic absolute magnitudes are as follows:

$$\delta\ \text{Cep}\quad M_{\text{pg}_{\text{med}}} = -2\overset{m}{.}70 \pm 0\overset{m}{.}04$$

$$\eta\ \text{Aql}\quad M_{\text{pg}_{\text{med}}} = 3\overset{m}{.}67 \pm 0\overset{m}{.}40$$

TABLE 5

Star	HD spectral type	Spectral types, Yerkes classification				M_V	Our values		M_V, trig	No. of spectrograms
		MKK [11]	MK [21]	Roman [27]	Gathier [28]	Gathier [28]	M_V	spectral type	[23]	
					Supergiants					
α Per	F 5	F 5 I b	F 5 I b	—	F 5+ II	—	$-4^m.0 \pm 0^m.3$	F 5 I b—II	—	4
γ Cyg	F 8	F 8 I b	F 8 I b	--	F 7 I b	—4.0	—5.9 ± 0.3	F 8 I b	—	3
β Aqr	G 0	G 0 I b	G 0 I b	--	G 0 I b	—3.5	—4.2 ± 0.4	G 0 I b—II	—	2
α Aqr	G 0	G 1 I b	G 2 I b	—	—	—	—5.0 ± 0.3	G 2 I b	—	2
β Dra	G 0	G 2 I b	G 2 II	—	—	—	—5.3 ± 0.2	G 2 I b	—	2
9 Peg	G 5	G 5 I b	G 5 I b	—	C 4 I b—II	—	—4.0 ± 0.4	G 5 I b—II	—	2
					Subgiants					
θ Cyg	F 5	—	—	F 5 IV$_s$	—	—	—	F 5 IV	$+3^m.7 \pm 0^m.2$	2
ξ Peg	F 5	F 6 III—IV	F 7 V	F 6 III—IV$_w$	—	—	+2.3 ± 0.4	F 6 III—IV	+2.7 ± 0.3	2
υ Peg	G 0	F 8 III	F 8 IV	F 8 III$_s$	F 7 II	0.0	+1.0 ± 0.4	GO III	+1.5 ± 0.5	2
ζ Her	G 0	G 0 IV	G 0 IV	G 0 IV$_w$	G 2 IV	+3.3	+3.3 ± 0.3	G 1 IV	+3.2 ± 0.1	2
μ Her	G 5	G 5 IV	G 5 IV	G 5 IV$_w$	—	—	—	G 5 IV	+3.6 ± 0.1	2
η Her	K 0	G 5 III	—	G 8 III—IV$_s$	—	—	+1.5 ± 0.4	G 5 III	+2.2 ± 0.2	2

Cepheids

Star	Phase	Our values				No. of spectrograms
		spectral type	M_V in phase	M_V median	M_{pg} median	
δ Cep	$0^P.50$	G 1 I b—II	$-3^m.27 \pm 0^m.24$	$-3^m.43$	$-2^m.72$	2
δ Cep	0.65	G 1 I b—II	—3.36 ± 0.26	—3.68	—2.94	2
δ Cep	0.74	F 9+ I b—II	—2.96 ± 0.30	—3.34	—2.61	3
δ Cep	0.89	F 9 I b—II	—3.33 ± 0.51	—3.27	—2.67	3
η Aqr	0.20	F 8 I b	—4.46 ± 0.40	—4.35	—3.67	3

3. Zero-Point of the Period-Luminosity Relation of Long-Period Cepheids

The values obtained for the median absolute magnitudes of Cepheids enable us to determine the zero-point of the period-luminosity relation, which determines the scale of distances in the universe; this accounts for the great interest in the zero point and the great amount of work on its determination.

In 1952-1954 it was shown that between the period-luminosity relations of short-period and long-period Cepheids there is a discontinuity at $1.^{m}5$-2^{m}. This observed fact has been interpreted in different ways.

Baade [30] states that the mean absolute magnitude of short-period Cepheids is $0.^{m}0$ and that the correction to the zero point of Shapley's period-luminosity relation [34] is $-1.^{m}5$ for long-period Cepheids.

Parenago [31] states that the mean absolute magnitude of short-period Cepheids is $+0.^{m}5$. This order of magnitude is obtained both from the proper motions and from the parameters of a subsystem of short-period Cepheids [28]. The correction to Shapley's zero-point is therefore the order of -1^{m}.

Kholopov [32] states that Shapley's zero point is correct and that the mean absolute magnitude of short-period Cepheids is of the order of $+2^{m}$.

Kukarkin [33] showed in 1949 that the period-luminosity relation can be represented by three formulas, one of which, for Cepheids with periods from 1 to 9 days, is the following:

$$M_{pg_{med}} = -0^{m}\!.74 - 1^{m}\!.67 \log P \tag{3}$$

Kukarkin obtained the correction $-0.^{m}5$ to Shapley's zero point for long-period Cepheids.

Using Formula (3) we obtained from the photographic median absolute magnitudes of δ Cep and η Aql the correction $-0.^{m}76 \pm 0.^{m}40$ to Kukarkin's zero point (the error includes the entire dispersion of absolute Cepheid magnitudes, which is $\pm 0.^{m}30$).

There are many recent determinations of the correction to the zero point, some of which have been considered by Parenago [31]. He determined the correction to the zero point from the mass-luminosity relation for supergiants and from the trigonometric parallaxes of Cepheids. He converted to his system the results obtained by Blaauw and Morgan for the zero point from the proper motions of 18 Cepheids, and by Kukarkin from the effect of galactic rotation on the radial velocities of Cepheids, and also used the investigation by Thackeray and Wesselink of short-period Cepheids in the Magellanic clouds.

TABLE 6

Method	Correction to Kukarkin's zero point	Correction to Shapley's zero point
Distance to center of Galaxy from radial velocities of Cepheids [39]	$-0.^{m}8 \pm 0.^{m}4$	$-1.^{m}3 \pm 0.^{m}4$
Proper motions of Cepheids [40]	− 0.8 ± 0.4	− 1.3 ± 0.4
Proper motions of Cepheids [41]	− 0.9 ± 0.5	− 1.4 ± 0.5
Short-period Cepheids in Magellanic Clouds [42]	− 0.8 ± 0.4	− 1.8 ± 0.4
Bright stars in globular clusters of Magellanic Clouds [43]	− 1.1 ± 0.4	− 2.1 ± 0.4
Integr. mag. of globular clusters of Magellanic Clouds [43]	− 0.8 ± 0.5	− 1.8 ± 0.5
Mass-luminosity relation for Magellanic Clouds [31]	− 0.7 ± 0.9	− 1.2 ± 0.9
Trig. parallaxes of Cepheids [31]	− 0.9 ± 1.6	− 1.4 ± 1.6
Abs. mag. of Cepheids from luminosity effect	− 0.7 ± 0.4	− 1.3 ± 0.4
Rel. bolom. mag. − radius and temp. effect	− 0.7 ± 0.4	− 1.2 ± 0.4

This summary of determinations of the zero point can be supplemented by a few additional values, converted to the system employed by Parenago (mean photographic absolute magnitude of short-period Cepheids $+0.^{m}5$, light absorption by Parenago's method, velocity of the sun with respect to the Cepheids 18.6 km/sec and distance to the center of the Galaxy from radio observations 8.2 kiloparsecs [35]).

Besides Blaauw and Morgan the zero point was determined from the proper motions of 18 Cepheids by A. Ia. Filin in 1953. But in calculating light absorption he exaggerated the dispersion of the absolute magnitudes of his Cepheids ($\pm 0.^{m}63$ instead of $\pm 0.^{m}44$ when calculated as indicated by Kukarkin [33]). The correction of this error and conversion to Parenago's system (according to Parenago [21] the reduction for the catalogues is $+0.^{m}25$) gives a correction to the zero point of $-0.^{m}9 \pm 0.^{m}5$ (mean error), which agrees within the limits of error with Blaauw and Morgan in the unit system (Table 6). Filin's correction to Kukarkin's zero point was originally $-1.^{m}81 \pm 0.^{m}44$ (rms error).

In 1954 H. Weaver determined the correction to Shapley's zero point from the discrepancy between the estimates of the sun's distance from the center of the Galaxy as determined from the radial velocities of globular clusters and Cepheids. In its original form the correction was $-1.^{m}56 \pm 0.^{m}30$ (mean error). When converted to Parenago's system the correction is of the same magnitude as by the preceeding methods.

A correction to the zero point was also obtained from the distance modulus to the Magellanic Clouds. Considering the new work of Thackeray and Wesselink on the apparent magnitude of short-period Cepheids in the Magellanic Clouds, we obtained for the distance modulus to the large Magellanic Cloud $18.^{m}1 \pm 0.^{m}4$ and to the small Magellanic Cloud $18.^{m}2 \pm 0.^{m}4$ (probable error), considering the mean photographic absolute magnitude of short-period Cepheids to be $+0.^{m}5$. Two additional methods for determining the distance modulus have been suggested by Shapley. By the first of these methods (comparison of the apparent magnitudes of bright stars in globular clusters of the Galaxy and Magellanic clouds) the distance modulus to the large Magellanic cloud is $18.^{m}5 + 0.^{m}4$ and to the small Magellanic cloud $18.^{m}6 \pm 0.^{m}4$ (in our system). By the second method (comparison of the integral stellar magnitudes of globular clusters in the Galaxy and Magellanic Clouds) the distance moduli are $18.^{m}2 \pm 0.^{m}5$ (large Magellanic cloud) and $18.^{m}2 \pm 0.^{m}5$ (small Magellanic cloud). The results are given with mean errors. The average distance moduli are identical for the large and small Magellanic clouds and for each method can be taken as the average for the large and small clouds.

Using the period-luminosity relation in apparent magnitudes for the Magellanic clouds as given by Kukarkin [33], we determined the zero point by each of the three methods. Opolski and Krawiecka [44] by determining the radius and estimating the effective temperatures of δ Cep and η Aql obtained their mean bolometric magnitudes and a correction of $-1.^{m}4$ to Shapley's zero point. Conversion to the more accurate temperature scale of Parenago [31] reduces the correction to $0.^{m}2$.

Table 6 is a list of ten determinations of the correction to the period-luminosity zero points of Kukarkin and of Shapley (1930), which were obtained from 1953 to 1956. This list does not include the estimates of the zero point from novae (Kopylov [36], Buscombe and Vaucouleurs [37]), uncertain consequences of the large real dispersion of maximum absolute magnitudes of novae and the estimate of the zero point from parameters of the subsystems of long-period Cepheids (Kopylov and Kumaigorodskaia [38]) which is apparently not free of systematic errors.

There is remarkably good agreement among the majority of corrections to Kukarkin's zero point (assuming the mean photographic absolute magnitude of short-period Cepheids to be $+0.^{m}5$). The same agreement is not found among corrections to Shapley's zero point (mean photographic absolute magnitude of short-period Cepheids $0.^{m}0$); the corrections to the zero point according to the short-period Cepheids in the Magellanic clouds are greater by $0.^{m}6$ in absolute value than by other methods (Table 6). Thus the most acceptable mean photographic absolute magnitude of short-period Cepheids is that of Parenago, and Parenago's view regarding the discontinuity between the mean absolute magnitudes of short-period and long-period Cepheids is closest to the truth.

The most probable correction to Kukarkin's zero point for the period-luminosity relation is $-0.^{m}83 \pm 0.^{m}04$ (rms error), which corresponds to multiplying Shapley's distance scale by the factor 1.8.

P. K. Shternberg State Astronomical Institute

Received October 1, 1956

LITERATURE CITED

[1] J. L. Greenstein, Proc. Natl. Acad. Sci. 26, 259 (1940).

[2] R. E. Wilson, Astrophys. J. 93, 212 (1941).

[3] P. C. Keenan and W. W. Morgan, Astrophysics (ed. J. A. Hynek),Chap. I (New York, 1951).

[4] R. E. Wilson, General Catalogue of Stars' Radial Velocities (Wash. Obs. Publ., 1953).

[5] M. W. Feast, A D. Thackeray and A. J. Wesselink, Mem. R. A. S., 67, part 2 (1955).

[6] P. P. Parenago, Trudy GAISh 20, 26 (1951).

[7] R. E. Wilson, Astrophys. J. 92, 170 (1940).

[8] O. A. Mel'nikov, Trudy GAO 64 (1950).

[9] K. F. Ogorodnikov, Astron. J. (USSR) 21, 43, 46 (1944).

[10] P. P. Parenago, Astron. J. (USSR) 22, 129 (1945).

[11] W. W. Morgan, P. C. Keenan and E. Kellman, An Atlas of Stellar Spectra (Yerkes Obs., Chicago, 1943).

[12] K. Bottlinger. Astron. Nachr. 212, 142 (1921).

[13] N. Rassmusson, Lund. Medd., Ser. II, 26, 28 (1921).

[14] H. Schloss,Z. Astrophys. 11, 117 (1935).

[15] W. Smart, and A. Ali, Monthly Notes 100, 560 (1940).

[16] R. Trumpler, Lick Obs. Bull. 14, 154 (1930).

[17] N. G. Roman and W. W. Morgan, Astrophys. J. 111, 426 (1950).

[18] M. K. V. Bappu, Monthly Notices 114, 687 (1954).

[19] O. J. Eggen, Astron. J. 60, 65 (1955).

[20] D. L. Harris, Astrophys. J. 123, 371 (1956).

[21] W. W. Morgan and J. L. Johnson, Astrophys. J. 117, 313 (1953).

[22] H. L. Johnson and W. A Hiltner, Astrophys. A. 123, 276 (1956).

[23] L. F. Jenkins, General Catalogue of Trigonometric Stellar Parallaxes (Yale Univ. Obs., New Haven, Conn., 1952).

[24] Publications Allegheny Obs., 10 (1956).

[25] W. A. Hiltner and R. C. Williams, Photometric Atlas of Stellar Spectra (Univ. of Michig. Press, Ann Arbor, 1946).

[26] W. J. Iwanowska, Wilno Obs. Bull., 18 (1937).

[27] N. G. Roman, Astrophys. J. 116, 122 (1952).

[28] P. J. Gathier, Rech. Astr. Obs. Utrecht., 13, part 2 (1955).

[29] O. J. Eggen, Astrophys. J. 113, 367; 114, 144 (1951).

[30] W. Baade, Trans. J. A. U. 8, 397 (1954).

[31] P. P. Parenago, Variable Stars (USSR) 10, 193 (1954).

[32] P. N. Kholopov, Astron. Tsirkuliar 148 (1954).

[33] B. V. Kukarkin, Peremen. Zvezd. 7, 57, 69 (1949).

[34] H. Shapley, Star Clusters, Harv. Mon., 2 (1930).

[35] P. P. Parenago, Astron. J. (USSR) 32, 226 (1955).

[36] I. M. Kopylov, Izv. Krim. Astr. Obs. 13, 76 (1955).

[37] W. Buscombe and G. Vaucouleurs, Astron. J. 60, 155 (1955).

[38] I. M. Kopylov and R. N. Kumaigorodskaia, Izv. Krim. Astr. Obs. 15, 169 (1955).

[39] H. Weaver, Astron. J. 59, 375 (1954).

[40] A. Blaauw and H. Morgan, Berlin. Astron. Nachr. 12, 95 (1954).

[41] A Ia. Filin, Biul. Stalinab. Obs. 10, 14 (1954).

[42] A Thackeray and A. Wesselink, Monthly Notices A. S. S. A. 13, 99 (1954).

[43] H. Shapley and V. Mc Kibben, Proc. Nat. Acad. Sci. 40, 1 (1954).

[44] A Opolski and J. Krawiecka, Contr. Wroclaw Obs. No. 11 (1956).

THE DISTRIBUTION OF ABSORBING MATTER AND THE SPIRAL STRUCTURE OF THE GALAXY

N. E. Kurochkin

The absorption of light in a heterogeneous medium is given in stellar magnitudes by $A(r)=m-m_0=1.086\kappa\int_{r_1}^{r_2}\rho(r)\,dr$, where κ is the absorption coefficient per unit density and $\rho(r)$ is the density. For small values of r_2-r_1 the density $\rho(r)$ may be taken as constant. The distribution of density in the galaxy may then be studied, without any additional assumptions, to within κ. Absorption studies have been carried out in this way by Kharadze [1] and van Rhijn [2]. The density distribution as a function of the z coordinate and the radius (the latter was obtained from the absorption data of van Rhijn for six Kapteyn areas with $b \le \pm 3°$ assuming spherical symmetry) may be represented by

$$D_z=24.38\cdot10^{-4}e^{-z/216.5};\qquad D_R=19.28.10^{-4}e^{-(R-7200)/1010}$$

where r and z are expressed in parsecs.

The deviation from the logarithmic law of distribution of density for low latitude directions agrees with the spiral structure observed in the solar vicinity, determined from the distribution of maxima of B-stars (according to Bergedorf Spektral Durchmusterung [11]) and the maxima in the distribution of neutral hydrogen (radio observations [7]). The spiral structure in the solar vicinity is very irregular but its general form can be deduced from the sum total of observations. More attention should be given to theoretical and experimental studies of physical properties of interstellar medium in order to determine κ and the zero point of the derived regularities.

Studies of galactic structure meet with very great difficulties because of the presence in the galaxy of light-absorbing medium. It is difficult to say whether it is the absorbing matter or the stars which are the more irregularly distributed over small regions of space. Undoubtedly, the visible structure of the stellar system, which is mainly determined by weak stars, is markedly distorted by absorbing matter. A detailed study of the general laws of distribution of absorbing matter should therefore precede studies of the distribution of stars. Systematic studies of absorption of light which were begun in the thirties have produced a number of papers on the subject. Kharadze [1] has given a sufficiently detailed review of this work. Among these papers only those are fundamental which give a sufficiently accurate and extensive series of observations of the color indices from which, knowing the spectral class of the stars, one can obtain the color excesses and full absorption in the photographic and visual regions. Undoubtedly, an accurate determination of the system of color indices is of major importance, and in considering the data of different authors differences in the systems must be taken into account. We shall show that the same treatment applied to results of different authors leads to strongly differing conclusions. This is especially the case in determinations of light absorption for weak objects, which correspond to those parts of the galaxy which are remote from the sun. The determination of color indices and spectra of weak stars is of importance if we wish to study general regularities in the distribution of dark matter. The distribution of absorbing matter fluctuates so much that absorption studies over small parts of the sky give no clue as to the general distribution laws. At the same time, and for the same reason, studies of weak absorbing objects over the whole of the sky is also pointless since it would only strengthen our ideas on the chaotic way in which dark matter is distributed. Detailed studies should only be carried out along those directions in which one would expect to discover general and most characteristic properties of the distribution of matter, and, particularly, where one can penetrate to large distances from the sun where near clouds of dark matter are absent. The existing data on the absorption of light and the galactic structure in the region 1-2 kiloparsecs from the sun may be used in a preliminary study of the distribution of absorbing matter, and this is done in the present paper. Accurate photoelectric data on color indices of weak stars in chosen areas, where one would expect structural properties of the galaxy to show up, and on visibility at large distances (the work on this is not very voluminous and requires powerful instrumentation) could be used to solve the problem of the distribution of absorbing matter sufficiently accurately and completely. The choice of such directions is bound up with the elucidation of the position of spiral arms in the solar region, and preliminary studies of the distribution of dust relative to spirals, stellar clouds and other stellar formations.

Method. The absorption of matter in a heterogeneous medium is given by

$$dE = -k(r)\,E\,dr, \tag{1}$$

where E is the intensity of light, k(r) the absorption coefficient depending on the distance r and, in general on the wavelength λ, although usually one finds the total photographic absorption on the international scale of stellar magnitudes and color indices which gives k(r) averaged over the given spectral region. Let us express k(r) in the form $k(r) = \kappa\rho(r)$, where κ is the absorption coefficient per unit density and depends only on the physical properties of the medium. If we study the distribution of matter over a large region of space then this assumes the independence of κ of r, which is true in the galaxy only to a first approximation. Differences are possible in the composition of absorbing matter at different distances from the galactic nucleus, at different distances from the plane of the galaxy and even in different dust clouds. As the various methods of studying interstellar matter develop it will be possible to bring out these differences.

Integrating the equation $dE = -\kappa\rho(r)E\,dr$ over a range of r we have

$$\ln\frac{E}{E_0} = -\varkappa\int_{r_1}^{r_2}\rho(r)\,dr. \tag{2}$$

Hence, multiplying by -2.5, and converting to logarithms to the base 10 we have

$$A(r) = m - m_0 = 1.086\varkappa\int_{r_1}^{r_2}\rho(r)\,dr. \tag{3}$$

If we have values of A(r) for different distances along the line of sight, the differences between them will be equal to absorption over the path difference $r_2 - r_1$. For small differences $r_2 - r_1$ the density may be taken as constant and equal to some mean value $\bar{\rho}(r)$ corresponding to the difference $r_2 - r_1$. In that case

$$\varkappa\bar{\rho}(r) = \frac{A(r_2) - A(r_1)}{1.086(r_2 - r_1)}.$$

We thus have the linear or volume density of the absorbing material, depending on the method of calculation of the absorption coefficient κ which can, for example, be expressed in stellar magnitudes per parsec or per unit density.

We thus have the possibility of measuring the density of absorbing matter to within the constant multiplier (or on the logarithmic scale, to within the additive constant). By determining κ by some method we can get an idea of the zero point of $\rho(r)$. Random density fluctuations noticeable even over large distances $r_2 - r_1$, of the order of 100 parsecs or more, allow us to average the values of $\rho(r)$ over distances of 200-300 parsecs, which is not reflected on the general regularities in the distribution where distances of a few hundreds of parsecs or kiloparsecs are significant. Naturally, the smaller the interval the more accurate are the observations, and the finer the structure shown, but it is debatable whether one can expect more than a confirmation of randomness in the density distribution from such observations. Only a comparison of the density of absorbing matter with the distribution of other matter, and studies of general regularities in the distribution of matter in the galaxy, can have any meaning. The above method frees us, to a considerable extent, from various preliminary assumptions as to the nature of the distribution of matter in space, and allows us to study densities along different directions in the galaxy. The determination of κ thus becomes an important problem. Here it is possible to carry out, in addition to a theoretical treatment, an experimental determination of the absorption coefficient of meteoritic matter in the form of dust consisting of particles 10^{-5} cm in diameter and suspended in some transparent liquid or gas. A comparison of such experimental data with theoretical deductions, and with observations, would enable us to establish the physical properties of absorbing matter quite reliably.

Distribution of absorbing matter along the z-coordinate. To determine the density of absorbing matter by the method described above we used the data of Kharadze [1] and van Rhijn [2]. In the latter work the data of Kharadze were used to start with, but were corrected using photoelectric and photometric observations of various authors [3-6]. The work of van Rhijn must therefore be given precedence. Observations of Kharadze and other authors contain determinations of color excesses for stars of early spectral classes in chosen Kapteyn areas between $l = 0°$ and $l = 143°$. Using the conversion factor $\gamma = 5.2$ (from Kharadze's data; this value was confirmed by van Rhijn for the international system of color indices in a summary of Leyden observations) selective absorption may be converted into full photographic absorption $A(r)$. Both Kharadze and van Rhijn give tables of $A(r)$ for different r. These tables were used in our calculations using Formula (4).

All SA areas were split into four groups with respect to the galactic latitude: I (l = 0-59°), II (l = 59-81°), III (l = 81-118°), IV (l = 118-143°). This division was carried out because of a correspondence with the possible distribution of spiral arms in the solar neighborhood to be discussed below. We must, however, admit that no real dependence on galactic longitude was found within the limits of the above groupings, nor from one group to another, since random fluctuations along different directions in the solar region are apparently much larger. All the same, some general regularities can be observed in regions which are more distant from the sun. Table 1 shows the grouping of the data. The magnitudes in high latitude areas 14, 15, 32 and 12, 13, 31 were averaged by van Rhijn. Mean values of l and b were taken for them (l = 75°, b = + 59.7° in the first case, and l = 119°, b = + 55.1° in the second case).

TABLE 1

	Kharadze	Van Rhijn
I SA	16, 33, 34, 35, 36, 37, 38,39, 40	38, 39, 40, 87, 110
II SA	6, 7, 14, 15, 17, 18, 19, 41, 42, 43	6, 7, 14—15—32, 17, 18, 19, 41, 42, 43
III SA	1, 2, 3, 4, 5, 8, 9, 13, 20, 21, 22, 32	1, 2, 3, 4, 5, 8, 9, 20, 21, 22
IV SA	10, 11, 12, 23, 24, 25, 26, 27, 28, 29, 30, 31	10, 11, 12—13—31, 23, 24, 25, 26

TABLE 2

Range	Group I					Group II					Group III					Group IV				
$\bar{z}$ (ps)	$\bar{z}$ (ps)	$\kappa\bar{\rho}\cdot 10^4$	$\log\kappa\bar{\rho}\cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$ (ps)	$\kappa\bar{\rho}\cdot 10^4$	$\log\kappa\bar{\rho}\cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$ (ps)	$\kappa\bar{\rho}\cdot 10^4$	$\log\kappa\bar{\rho}\cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$ (ps)	$\kappa\bar{\rho}\cdot 10^4$	$\log\kappa\bar{\rho}\cdot 10^4$	n	$\bar{r}_{xy}$
0—10	0.8	22.48	1.352	10	640	4.5	19.01	1.279	6	158	5.4	59.40	1.774	5	150	1.1	30.00	1.477	13	800
10—25	19.0	17.05	1.232	2	98	16.6	31.50	1.498	10	439	16.0	36.50	1.562	10	194	22.0	16.38	1.214	5	86
25—50	39.8	17.77	1.250	6	85	36.4	21.85	1.339	11	443	39.2	10.70	1.029	14	399	38.2	19.21	1.284	9	132
50—100	78.7	5.53	0.743	6	267	73.9	16.08	1.206	11	366	73.0	10.70	1.029	16	745	77.3	13.03	1.114	10	284
100—200	132.6	3.56	0.551	12	353	148.7	11.01	1.042	17	656	150.7	12.30	1.090	16	424	151.0	5.13	0.710	20	535
200—300	244.1	5.98	0.777	9	773	251.4	5.68	0.754	8	597	245.4	5.17	0.714	10	597	248.1	3.03	0.481	11	796
300—400	340.8	3.29	0.517	7	674	350.0	3.49	0.543	8	952	349.4	10.39	1.017	10	940	357.3	9.09	0.959	9	658
400—500	434.3	7.56	0.878	4	1047	450.3	1.08	0.033	6	1351	448.8	5.88	0.769	4	788	463.0	9.06	0.957	4	775
500—600	529.0	4.61	0.664	4	544	508.3	6.47	0.811	3	663	555.8	6.82	0.834	5	898	548.5	6.49	0.812	6	667
600—700	642.0	3.92	0.593	4	831	638.5	4.76	0.678	2	645	645.0	0.92	$\bar{1}.964$	2	1258	652.5	5.14	0.711	4	945
700—800	745.0	4.19	0.622	4	633	752.0	8.12	0.910	2	674	726.0	4.45	0.648	2	982	747.7	4.86	0.687	3	524
800—900	849.7	0.51	$\bar{1}.708$	3	1037	816.0	4.61	0.664	1	670	843.5	2.46	0.391	2	462	858.7	4.30	0.633	3	925
900—1000	942.0	4.61	0.664	1	566	922	0.00	—	1	600	903.0	1.53	0.185	1	1005	953.0	2.30	0.362	1	550
1000—200	1084.0	2.27	0.356	5	938	1098	1.64	0.215	3	843	1065.0	5.79	0.763	3	780	1074.7	2.66	0.425	3	652
1200—400	1274.0	5.28	0.723	3	801	1358	2.69	0.430	2	1054	1268.0	3.17	0.501	3	919	1281.0	3.59	0.555	3	779
1400—600	1527.5	5.22	0.718	2	788	—	—	—	—	—	1468.7	0.77	$\bar{1}.886$	3	1055	1450.5	3.45	0.538	2	895
1600—800	1800.0	7.38	0.868	1	1081	—	—	—	—	—	1678.0	2.76	0.441	1	1260	1662.5	0.76	$\bar{1}.881$	2	1022

Table 2 and Figure 1 show statistical results along the z-coordinate, based on Kharadze's data. In the first column are shown the values of z over which the average of $\kappa\rho$ was taken, in the second column – the mean value of z, next – the corresponding values of $\overline{\kappa\rho}$ and their logarithms, the number of points n, and the mean distance from the sun in the galactic plane r_{xy}. Figure 1 shows log $\overline{\kappa\rho}$ as a function of z. As can be seen, at small values of z the fall in the density may be quite well represented by the barometric law. At z = 100-200 parsecs a pronounced slow-down in the fall of the density occurs, and from this region onwards the points may be fitted approximately with another exponential having a smaller slope. In the absence of extensive and reliable data corresponding to larger values of z, one observes a strong scatter in the values of log $\overline{\kappa\rho}$ but, all the same, the discontinuity at z = 100-200 parsecs may be inferred from the general behavior of the distribution based on Kharadze's data. The following Formulas (5) represent the data given in Table 2 and Figure 1 ($D = \overline{\kappa\rho}$).

$$\begin{aligned} &\text{I. } D = 22.39\cdot10^{-4}e^{-z/72.5}\ (z < 110\ \text{ps}); && \text{III. } D = 54.90\cdot10^{-4}e^{-z/55.6}\ (z < 95\ \text{ps}); \\ &\phantom{\text{I. }} D = 4.90\cdot10^{-4}e^{-z/.700}\ (z > 110\ \text{ps}); && \phantom{\text{III. }} D = 12.30\cdot10^{-4}e^{-z/768}\ (z > 95\ \text{ps}); \\ &\text{II. } D = 29.17\cdot10^{-4}e^{-z/149.5}\ (z < 250\ \text{ps}); && \text{IV. } D = 30.00\cdot10^{-4}e^{-z/85.5}\ (z < 150\ \text{ps}); \\ &\phantom{\text{II. }} D = 6.53\cdot10^{-4}e^{-z/1665}\ (z > 250\ \text{ps}); && \phantom{\text{IV. }} D = 5.50\cdot10^{-4}e^{-z/3125}\ (z > 150\ \text{ps}) \end{aligned} \tag{5}$$

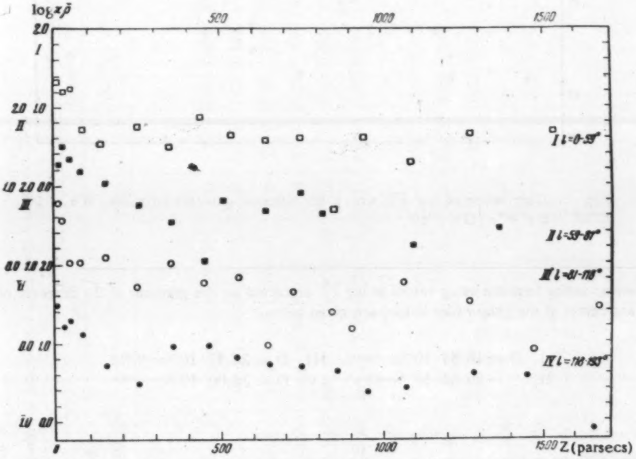

Fig. 1. Distribution of log $\overline{\kappa\rho}$ with z (from the data of Kharadze).

It is clear from the above that an application of the method of Parenago [7, 8], which assumes a single exponential law for all values of $\underline{z}$, gives an average value of the gradient of the distribution over all $\underline{z}$, while with the corresponding grouping with respect to the galactic latitude one should obtain the dependence of the gradient on $\underline{b}$. Such a dependence was found by Kochlashvili and Toronzhadze [9], who used a somewhat modified form of Parenago's method. To verify this, we split the material in group IV into three parts with respect to the latitude ($b < 20°$, $20° \leq b \leq 40°$, and $b > 40°$) and obtained the values of $\overline{\kappa\rho}$ and $\overline{z}$ for each subgroup. Because of the subsivision of the data, one now has a larger scatter of points, but as is seen from Figure 2, no connection between the density gradient as a function of $\underline{z}$, and the latitude $\underline{b}$, exists. In what follows, no such grouping of the material was found necessary. An attempt to discover the dependence of the characteristics of the distribution of density on the distance from the sun by suitable grouping of the data was also unsuccessful.

The problem of the existance of changes in the density gradient at large values of $\underline{z}$ had to be examined. Van Rhijn's data, which are undoubtedly the more reliable, were used in this. Table 3 and Figure 3 show the corresponding results of calculations. As can be verified, the dependence of log $\overline{\kappa\rho}$ on $\underline{z}$ at different longitudes is sufficiently well represented by straight lines. The distributions can be represented by the following equations:

$$\begin{aligned} &\text{I. } D = 18.41 \cdot 10^{-4} e^{-z/249.5}; \quad \text{III. } D = 21.88 \cdot 10^{-4} e^{-z/170.3}; \\ &\text{II. } D = 27.54 \cdot 10^{-4} e^{-z/165.8}; \quad \text{IV. } D = 20.89 \cdot 10^{-4} e^{-z/221.6}. \end{aligned} \tag{6}$$

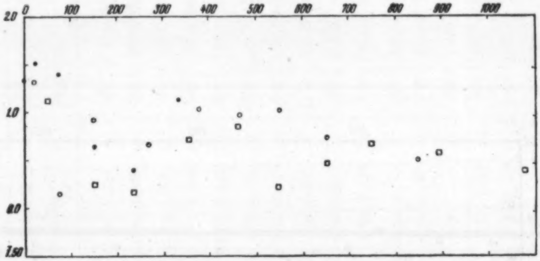

Fig. 2. Distribution of log $\overline{\kappa\rho}$ with $\underline{z}$ for different galactic latitudes. ● $b < 20°$; ○ $20° < b < 40°$; ☐ $b > 40°$.

The corresponding formulas using values of log $\overline{\kappa\rho}$ corrected for the gradient of the decrease of density towards the anticenter of the galaxy (see below) are given below.

$$\begin{aligned} &\text{I. } D = 18.84 \cdot 10^{-4} e^{-z/205.3}. \quad \text{III. } D = 23.17 \cdot 10^{-4} e^{-z/223.8} \\ &\text{II. } D = 29.51 \cdot 10^{-4} e^{-z/154.0}. \quad \text{IV. } D = 26.00 \cdot 10^{-4} e^{-z/278.4}. \end{aligned} \tag{7}$$

It is very difficult to discover a systematic dependence of the distribution of density on direction. For this reason data for four different directions were averaged and the magnitudes obtained (Table 4) were treated using the method of least squares. The final results in Table 4 (last two columns) contain a correction for the mean

TABLE 3

Distribution of Absorbing Matter from Van Rhijn's Data

Range of Z(parsecs)	Group I					Group II					Group III					Group IV				
	$\bar{z}$	$\bar{\kappa}_\rho \cdot 10^4$	$\log \bar{\kappa}_\rho \cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$	$\bar{\kappa}_\rho \cdot 10^4$	$\log \bar{\kappa}_\rho \cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$	$\bar{\kappa}_\rho \cdot 10^4$	$\log \bar{\kappa}_\rho \cdot 10^4$	n	$\bar{r}_{xy}$	$\bar{z}$	$\bar{\kappa}_\rho \cdot 10^4$	$\log \bar{\kappa}_\rho \cdot 10^4$	n	$\bar{r}_{xy}$
0—10	2.3	21.34	1.329	39	336	3.8	22.92	1.360	15	239	2.4	19.26	1.284	39	241	1.8	27.81	1.460	33	207
10—25	17.6	17.60	1.246	8	89	15.0	31.02	1.492	14	154	17.4	18.15	1.259	20	202	15.4	18.17	1.260	14	243
25—50	36.4	15.87	1.200	9	170	35.1	19.99	1.301	18	168	36.4	16.84	1.226	25	313	35.8	15.76	1.197	17	167
50—100	73.0	11.99	1.079	9	339	72.8	11.27	1.053	21	290	69.6	14.16	1.151	24	232	75.2	14.50	1.161	18	303
100—200	141.0	11.49	1.061	8	612	140.6	8.08	0.907	18	437	139.0	12.57	1.099	24	397	152.1	9.89	0.995	15	483
200—300	241.0	7.50	0.875	1	711	241.9	5.54	0.744	8	497	238.6	6.06	0.782	12	628	243.8	6.67	0.824	6	716
300—400	344.0	5.68	0.754	2	1029	344.8	4.54	0.657	5	788	336.8	2.73	0.436	5	672	341.0	7.38	0.868	4	520
400—500	—	—	—	—	—	453.7	0.28	$\bar{1}.447$	3	766	425.3	1.40	0.146	3	644	468.0	2.88	0.459	2	599
500—600	—	—	—	—	—	549.5	0.65	$\bar{1}.813$	2	540	536.0	0.77	$\bar{1}.886$	3	810	556.0	1.32	0.121	1	1070
600—700	—	—	—	—	—	658	0.54	$\bar{1}.732$	1	384	662.0	1.16	0.064	3	861	625.0	1.07	0.029	1	436
700—800	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
800—900	—	—	—	—	—	827	0.41	$\bar{1}.613$	1	483	821.0	0.16	$\bar{1}.204$	2	1088	—	—	—	—	—
900—1000	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

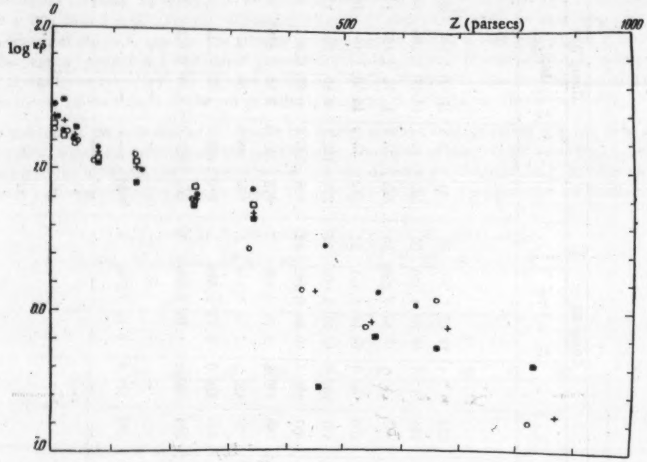

Fig. 3. Distribution of absorbing matter with z using van Rhijn's data. □) I l = 0-59°; ■) II l = 59-81°; ○) III l = 81-118°; ●) IV l = 118-143°; +) mean.

distance from the sun in the plane of the galaxy, which depends on the density gradient along the direction towards the center of the galaxy. The following formula represents the distribution of absorbing matter as a function of z in a final form:

$$D = \underset{\pm 0.68}{24.38} \cdot 10^{-4} e^{-z/216.5 \pm 15.1} \tag{8}$$

From a consideration of diagrams based on van Rhijn's data, the starting point of the theory of absorption developed by P. P. Parenago [7, 8] is confirmed, provided that the observational material is not subject to systematic errors. The numerical value of the height of the "homogeneous atmosphere" turns out to be twice as big as that assumed by Parenago. This means that the decrease in density takes place at a slower rate than was assumed up to now. Using Formula (4) we can consider the mean absorption coefficient in the photographic region in the galactic plane (a_0). With $r_2 - r_1 = 1$, $A(r_2) - A(r_1) = a_0$ and $a_0 = 1.086\,\overline{\kappa\rho}$. According to van Rhijn's data $\overline{\kappa\rho}_0 = 24.4 \cdot 10^{-4}$ stellar magnitudes per parsec, which corresponds to $2^{\text{m}}.65$ per kiloparsec. According to Kharadze $\overline{\kappa\rho}_0 = 30.2 \cdot 10^{-4}$, i. e., $3^{\text{m}}.28$ per kiloparsec. Kochlashvili and Toronzhadze obtained $3^{\text{m}}.95$ for all areas, and $3^{\text{m}}.4$ with the exclusion of areas in the zone of the galactic equator.

TABLE 4

Distribution of Densities with z According to Van Rhijn's Data (Average d for all Directions)

δ

$\bar{z}$	$\overline{\varkappa\rho}\cdot 10^4$	$\log \overline{\varkappa\rho}\cdot 10^4$	n	$\bar{r}_{xy}$	Corrected	
					$\overline{\varkappa\rho}\cdot 10^4$	$\log \overline{\varkappa\rho}\cdot 10^4$
2.4	22.58	1.354	126	261	24.95	1.397
16.3	21.29	1.328	56	184	23.61	1.373
35.9	17.27	1.237	69	222	19.39	1.288
72.4	13.13	1.118	72	280	15.21	1.182
142.7	10.58	1.025	65	454	13.26	1.122
240.8	6.09	0.785	27	612	9.08	0.958
341.2	4.83	0.684	16	714	6.68	0.825
446.6	1.35	0.130	8	678	2.18	0.338
543.8	0.82	$\bar{1}.914$	6	763	1.53	0.185
679.3	0.73	$\bar{1}.863$	7	758	1.19	0.076
865.2	0.18	$\bar{1}.255$	4	837	0.29	$\bar{1}.462$

Distribution of absorbing matter in the plane of the galaxy. Six areas (SA 110, 87, 19, 8, 9, 24) with galactic latitudes 0, 17, 80, 92, 107, 128° are correspondingly disposed within ± 3.1° of the galactic latitude. The values of $\overline{\kappa\rho}$ according to van Rhijn's data were reduced to the galactic plane using Equations (6). Knowing the distance of the sun from the center of the galaxy R_0 (assumed equal to 7.2 kiloparsecs), one can calculate for different $r_{xy} = (r_2 + r_1)/2$, corresponding to the mean densities, the distance of each area from the center of the galaxy using

$$R_{xy}^2 = r_{xy}^2 + R_0^2 - 2r_{xy}R_0\cos(l - l_0). \tag{9}$$

Here $l - l_0$ is the difference of longitudes of the given direction and the galactic center ($l = -32°$). Assuming circular symmetry in the distribution of matter one can obtain mean values of $\overline{\kappa\rho}$ for equal distances from the center. Table 5 and Figure 4 give results of these calculations. The density of absorbing matter as a function of radius may be represented by

$$D_R = 19.28\cdot 10^{-4}e^{-(R_{xy}-7200)/1010}. \tag{10}$$

The final distribution of density as a function of z was corrected using this formula and is given in Table 4.

Using the law found above, one can calculate the density of absorbing matter at the center of the galaxy. With R_0 = 7.2 kiloparsecs the absorption per parsec is $a_0 = 2^m.83$ which is obviously an overestimate. It is possible that the distance of the sun from the galactic center is much too large and it is also possible that our law is not sufficiently accurate for extended extrapolation. If one assumes R_0 = 5 kiloparsecs then a_0 is $0^m.27$ and with R_0 = 4 kiloparsec a_0 = 0.11. These magnitudes are more reasonable although for the full absorption in the plane of the galaxy one still obtains values which are somewhat high: 5717, 549 and 222^m respectively.

Undoubtedly in any determination of the total mass of absorbing matter in the galaxy, one must take into account the decrease in density both with the z-coordinate and the radius. An integration over an infinite ellipsoid of rotation using Expressions (8) and (10) gives

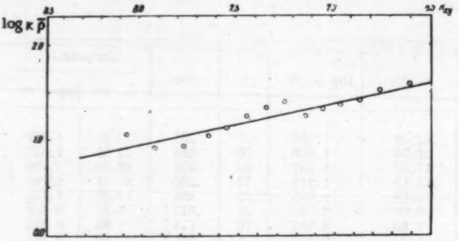

Fig. 4. Distribution of absorbing matter along the radius vector in the plane of the galaxy (according to van Rhijn's data). R_{xy} is the distance from the galactic center.

TABLE 5

Distribution of density as a Function Distance from the Galactic Center

Range of R_{xy}	$\kappa\bar{\rho}\cdot 10^4$	$\bar{R}_{xy}$	n	$\log \kappa\bar{\rho}\cdot 10^4$	Range of R_{xy}	$\kappa\bar{\rho}\cdot 10^4$	$\bar{R}_{xy}$	n	$\log \kappa\bar{\rho}\cdot 10^4$
6.5—6.7	39.00	6.602	2	1.591	7.3—7.4	21.86	7.347	12	1.339
6.7—6.8	32.85	6.760	2	1.517	7.4—7.5	17.95	7.455	7	1.254
6.8—6.9	25.77	6.864	3	1.411	7.5—7.6	13.60	7.551	4	1.134
6.9—7.0	23.70	6.966	3	1.375	7.6—7.7	11.22	7.646	4	1.050
7.0 - 7.1	21.18	7.055	6	1.326	7.7—7.8	8.77	7.779	3	0.943
7.1—7.2	18.20	7.139	8	1.260	7.8—8.0	8.27	7.931	3	0.918
7.2—7.3	25.20	7.256	19	1.401	8.0—8.2	12.00	8.078	2	1.079

$$M = 2\pi\beta_z\beta_R^2\rho_0, \tag{11}$$

where ρ_0 is the density at the galactic center. If one assumes that the smallest possible value of the density of absorbing matter in the solar vicinity is $\rho = 10^{-25}$ g/cc, the density at the galactic center turns out to be $\rho = 1.32\cdot 10^{-22}$ g/cc. using Formula (10) in which ρ is taken instead of $\overline{\kappa\rho_0}$ and the gradient β_R remains constant. Substituting the values of ρ_0 and β_R, β_Z from Equations (8) and (10) into (11) we find that the total mass of absorbing matter in the galaxy is

$$M = 6.28\cdot 216\cdot 1010^3\cdot 1.32\cdot 10^{-22}\cdot$$
$$\cdot(3.08\cdot 10^{18})^3 g \sim 5\cdot 10^{41} g \approx 5\cdot 10^8 m_\odot$$

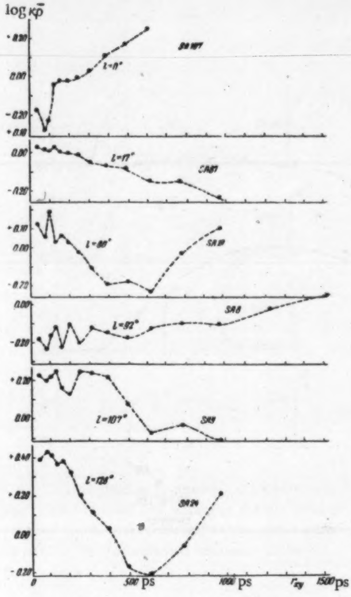

Fig. 5. Densities $\kappa\bar{\rho}$ for various direction in the Galactic plane [the excess $\Delta \log \kappa\bar{\rho}$ above the values predicted using Formula (10)].

A larger value is obtained if dispersion is taken into account. As a final value $10^9 m_\odot$ may be taken as the estimate of the mass of dust in the galaxy. This is higher by one order than the value accepted so far (Fesenkov's estimate is $10^8 m_\odot$ [10]).

It is possible that at some distance from the galactic center the rate of increase in the density decreases and as a result our estimate is too high.

Spiral structure of the galaxy in the solar vicinity. It is best to begin the study of the disposition of spiral arms with a study of the distribution of stars of early spectral classes, especially B and A-stars which are sufficiently numerous, can be observed at large distances, and have a tendency to concentrate in spiral arms and thus bring out the spiral structure of other galaxies. The distribution of other objects can be compared with the distribution of B-stars in order to establish their relation to the spiral arms. At the same time, studies of spiral arms in our galaxy are made rather difficult by the presence of major fluctuations in the distribution of B-stars.

A preliminary study of the distribution of B-stars was made using the Bergedorf Spektral Durchmusterung [11] which gives star counts for a given stellar magnitude for different Kapteyn areas. The simple formula

$$A(m) = 1.015 \cdot 10^{-4} \omega D(r) (r_2^3 - r_1^3) \qquad (12)$$

can then be used to study stellar density to within a constant coefficient (which depends on the dispersion of absolute magnitudes and luminosity function), at least for areas in low galactic latitudes. Here ω is the solid angle; r_1, r_2 are the distances corresponding to stars with some mean absolute magnitude (in our case it was assumed that for B0 — B5 stars $M = -3^m.4$. and absorption was taken into account using Parenago's tables [8]); A(m) is the number of stars of stellar magnitude between m and m + 1; D(r) is the stellar density. In order to remove the effect of factors not taken into account in (12), it is best to study not the densities themselves, which for the above reason may be subject to fluctuations but the maxima in the density distribution along the line of sight. It is known that maxima in the distribution of interstellar hydrogen obtained from radio observations [12] show a tendency to distribute themselves along certain curves which mark the spiral structure of the galaxy. By plotting the maxima in the distribution of B-stars it is possible to compare the distribution of B-stars and interstellar hydrogen. In Figure 6 the dotted line shows the positions of maxima in the hydrogen densities, the open circles indicate the positions of maxima in the distribution of B-stars, and open squares give the positions of regions of of O-associations which are also the clearest maxima of distribution of O- and B-stars. The number of areas employed was found to be too small for a study of the distribution of absorbing matter by this method. However, a comparison of Figures 6 and 5, showing the distribution of absorbing matter along the line of sight for the six areas which were investigated (in low galactic latitudes), allows us to draw certain conclusions as to the distribution of absorbing matter relative to the spirals in the galaxy.

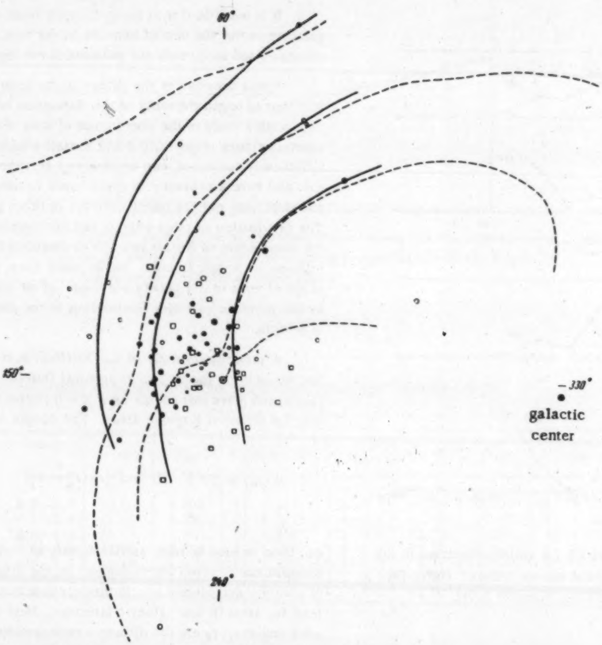

Fig. 6. Distribution of maxima of stellar densities of B-stars and the densities of interstellar hydrogen. Notation: dotted line – distribution lines of maxima in the density of hydrogen (after Parenago), ● and ○) maxima of B-stars (open circles are less reliable), □) areas of 0-associations according to Markarian [13]; continuous line – possible position of spiral arms.

It follows from the above graphs that two spiral arms are brought out in the neighborhood of the sun by the distribution of B-stars: one of these approaches the sun to a distance of about 1000 parsecs and is directed towards the anticenter of the galaxy, and the other approaches the sun very closely (200-300 parsecs) from the direction towards the galactic center. Maxima in hydrogen density confirm the tendency of B-stars to distribute themselves along these lines and to coincide with them in direction, although absolute coincidence is not observed. In spite of the errors involved in the determination of distances to B-stars and to hydrogen maxima (in the first case absorption was taken into account using Parenago's data and with β_Z = 100 parsecs, $a_0 = 3^m.5$ per kiloparsec which could lead to systematic errors in the distances, and in the second case, possible errors may be due to differential rotation of interstellar gas) the distribution of interstellar hydrogen may be much more "diluted" than the distribution of B-stars. If the spiral structure does not exist in the neighborhood of the sun, the hydrogen may "fill in" vacancies in the distribution of B-stars and penetrate into areas where B-stars are not observed. Undoubtedly, only a simultaneous study of B-stars and the interstellar medium can finally provide a solution to the problem of the location of spiral arms in the galaxy.

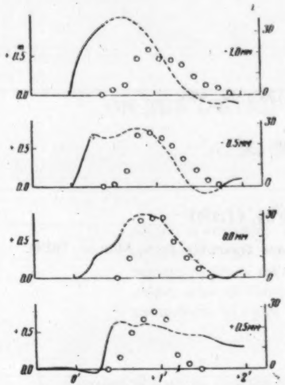

Fig. 7. Distribution of absorption (continuous and dotted curves) and luminous matter (photographic luminosity) (circles) in the spiral arms of NGC 5194 (for different sections) (after Holmberg).

Certain regularities are noticeable in the distribution of absorbing matter relative to the outlined spiral arms. The distribution of absorbing matter along some directions from the sun is plotted in Figure 5 (values of $\Delta \log \overline{\kappa\rho}$). Anticentral directions (l = 128, 107°) have a general tendency, which is at once noticeable, of giving large absorption in the neighborhood of the sun, and then a slow fall-off reaching a minimum (at distances of 500-700 parsecs) at places where one would expect a gap between two arms. It follows that the absorbing matter which accompanies spirals is more "smeared out" in space. Dark matter wraps itself round the spirals. This is confirmed by the distribution of along l = 0° where the line of sight enters into the arm nearer to us. Here the density increases as the line of sight enters the spiral (the data are free of the effect of the general density gradient along the radius). The data for other directions confirm this. Whenever the line of sight passes through the gap between the arms (l = 17°) the density slowly decreases (depending on the distance from the inner arm), and whenever it approaches the inner arm to a distance of 600-1000 parsecs an increase in the density is observed. Thus the distribution of the excess density of absorbing matter above the regular decrease along the radius vector confirms the general picture of the disposition of spirals in the neighborhood of the sun. In this connection one must mention the work of Holmberg [14] who studied absorption of light in the arm of the spiral galaxy NGC 5194 whose spiral screens the irregular galaxy NGC 5195. As can be seen from Figure 7, which was taken from [14], Holmberg found that the distributions of luminous and absorbing matter do not coincide with each other, the maximum of the distribution of dark matter is shifted relative to the arm towards the center of NGC 5194, and, finally, the absorbing matter occupies a larger region of space than the spiral which consists of stars. An analogous picture is apparently observed in our galaxy. Such a distribution of absorbing matter may be explained by radiation pressure which acts on the absorbing dust particles, the pressure being due to gigantic hot stars which fill the spiral arms of the galaxy. A random distribution of stars and interstellar matter is superimposed on the spiral structure and to a considerable extent masks it. For this reason, we cannot expect a very clear picture of the location of spiral arms in the solar neighborhood but should try to extract it from a chaos of randomness. One should not reject the possibility of the existence of spiral arms in the solar neighborhood solely because it is only brought out by the general tendency in the distribution of B-stars, interstellar hydrogen, and dark matter.

Spirals may turn out to be connected by "bridges" consisting of random clouds of stars and interstellar matter and located between the arms. Such a "bridge" may possibly be destroying the regular behavior of maxima of B-stars and interstellar hydrogen in directions from the sun towards l = 60°. In order to obtain a detailed picture of the location of spiral arms in the neighborhood of the sun one requires, undoubtedly, not only fuller data on the distribution of B-stars, but also the distributions of stars of other types, above all stars of spectral class A, which are more numerous and, with sufficiently powerful instruments, can be seen up to distances of 2-3 kiloparsecs.

Shternberg State Astronomical Institute

Received June 20, 1956

LITERATURE CITED*

[1] E. K. Kharadze, Abastuman, Astrophys. Observ. Bull. 1952 No. 12

[2] P. J. van Rhijn, Publ. Kapteyn Astron. Lab. 1955, No. 57.

[3] T. Elvius, Stockholms Observatory Ann. 16, 4, 5 (1955).

[4] F. H. Seares, Photovisual Magnitudes in the Selected Areas.

[5] Th. Walraven and A. D. Fokker, Bull. Astron. (Netherland, 1952), No. 435.

[6] F. H. Seares and M. C. Joyner, Astrophys. J. 98, 286 (1943).

[7] P. P. Parenago, Astron. J (USSR) 17, 4 (1940).

[8] P. P. Parenago, Astron. J. (USSR) 22, 3 (1945).

[9] T. A. Kochlashvili and A. F. Toronzhadze. Astron. J. (USSR) 31, 4 (1954).

[10] P. P. Parenago, Course of Stellar Astronomy* (State Tech. and Theor. Lit. Press, Moscow, 1954).

[11] Bergedorf Spektral Durchmusterung, Vol. 1-5.

[12] P. P. Parenago, Astrons J. (USSR) 32, 3 (1955).

[13] B. E. Markarian, Proc. Acad. Sci. Arm. SSR 15, 1 (1952).

[14] E. Holmberg, Lund. Medd. I, 170 (1950).

---

*In Russian.

ON THE DISTRIBUTION OF MASS IN THE GALAXY

A. M. Mikisha and F. A. Tsitsin

Lindblad's method for determining the relative mass of the galactic nucleus is analyzed. It is shown that a representation of the distribution of mass in the galaxy by a uniform model leads to more than a tenfold overestimation of the relative mass of the nucleus. A model of the galaxy is set up which takes into account the variation of flattening of the surfaces of equal density with the distance from the galactic center.

1. A method of studying the mass distribution in the galaxy has been given by Lindblad [1]. Themethod was used by a number of authors (Oort, Titus, Chandrasekhar and others), and is described, in its most highly developed form, by Parenago [2] in his "Course of Stellar Astronomy."

Since neither "solid rotation" (uniform spheroid, force in the galactic plane directly proportional to the distance from the center) nor "Keplerian rotation" (all mass concentrated at the center, force in the galactic plane inversely proportional to the square of the distance from the center) can be associated with the galaxy, Parenago [2] concludes that "in reality, part of the mass is concentrated at the galactic center and the rest is distributed throughout the galaxy (in the first approximation, uniformly)."

In that case ". . . one can try (our italics, A. M., F. Ts.) the following form of the force of attraction as a function of distance "

$$F = F_1 + F_2 = \alpha R^{-2} + \beta R. \tag{1}$$

(cf. [2] Section 34).

It must, however, be pointed out that this form cannot be taken as empirical. Since, in the first approximation, a uniform spheroid with a nucleus is taken as the model, the force on a point mass in the equatorial plane can be obtained quite accurately, and is given by Equation (1). Thus Equation (1) is introduced not because one can associate with the galaxy some sort of intermediate state of rotation between the "solid" and the "Keplerian" but because of the particular choice of a model.

In later calculations Lindblad and, following him, Parenago too, depart from the adopted model and,in fact,calculate not the mass of a uniform spheroid the the mass of a uniform flat disc.

A correct formulation of this problem (i. e., the determination of the mass of a uniform spheroid) is given by us below.

Equation (1) has often been criticized (Bottlinger [3], Parenago [2]). The main point of this criticism is the following. According to Equation (1) the force of attraction is inversely proportional to R^2 for small values of R and directly proportional to R for large values. This is contradicted by observational data which show that at small distances from the center the force is proportional to R. At large distances the force should, of course, fall off as R^{-2}. For this reason Equation (1) appears to be unsuitable.

This criticism is unsound since it is based on an erroneous interpretation of Equation (1).

Actually, Equation (1) gives an expression for the force acting on a point mass in the equatorial plane of a uniform spheroid with a nucleus. However, in its derivation, it was assumed that the nucleus attracts the point mass as a Newtonian center and, consequently, the point is situated outside the nucleus. On the other hand, the attraction due to the uniform spheroid was calculated for a point inside it. Thus Equation (1) is only true for points lying outside the nucleus but inside the spheroid and R cannot be made to tend to zero or infinity.

Rejecting Equation (1) Bottlinger [3] proposed the following expression for the force of attraction in the galactic plane:

$$F(R) = \frac{\alpha R}{1+\beta R^3}. \tag{2}$$

P. P. Parenago criticizes this expression and writes: ". . . this ingenious expression has no mechanical basis " [2].

However, it is easy to show the mechanical significance of Equation (2). For example,* in the case of a spherically symmetrical system, Equation (2) results in consequence of the following density distribution:

$$\rho(R) = \frac{3\alpha}{4\pi G}\frac{1}{(1+\beta R^3)^2}.$$

It would be interesting to consider a nonspherical distribution of mass giving rise to the law of force given by (2).

In spite of the fact that Bottlinger's criticism of Equation (1) is based on a misunderstanding, this form of the law of force must still be rejected since, as will be shown below, it corresponds to a model whose adoption leads to an overestimation of the relative mass of the galactic nucleus.

The immediate cause of our investigation was P. P. Parenago's statement in which he formulates the main result brought out by the above method of determining the mass of the nucleus and the whole galaxy: "Essentially, by determining $\mathfrak{M}_1$ (mass of nucleus, A. M., F. Ts.) one determines the mass distributed through galaxy within the distance between the sun and the center of the galaxy (this would be exact in a spherically symmetrical galaxy)" ([2] Section 34). Again, in another work, Parenago states: ". . .using Equation (12) (which corresponds to our Equation (1), A. M., F. Ts.) we determine not the mass of the central nucleus but the approximate mass contained within the solar distance (this would be true exactly in the case of a spherical galaxy)" [4].

However, it is easy to see that what is actually determined is the whole mass within the solar distance, but only if the law of force $F = \alpha R^{-2}$ holds (this is exact in the case of a spherical distribution of matter). Consequently, an analogous statement about the law of force given by Equation (1) is incorrect. Hence it follows that the large magnitude of the relative mass of the nucleus is largely due to the adoption of a uniform model of the galaxy.

To substantiate our point of view we shall consider the consequences of the adoption of a uniform model for nonuniform initial distribution with spherically symmetrical and spheroidally symmetrical distribution of matter. This method has the advantage that the initial distribution in each case is known, which is not the case when a model approximating a real galaxy is set up.

Before proceeding any further we must note the following. The initial distribution is fully determined by a finite number of parameters, all other parameters being expressible in terms of them. We shall take the numerical values of the parameters of the initial distribution to be equal to the numerical values of the corresponding parameters of the real galaxy. A comparison of the remaining parameters with the corresponding parameters of the real galaxy will, in addition to the main result, allow us to estimate how closely the initial distribution chosen by us corresponds to the real galaxy.

---

*We are indebted to I. L. Genkin for this idea.

2. Let δ be the density of the nucleus, $\underline{r}$ the radius of the nucleus, R_0 the distance of the sun from the galactic center, R_G the galactic radius, G the gravitational constant, $\mathfrak{M}_1$ the mass of the nucleus, M the mass contained in a sphere of radius R_0 $\mathfrak{M}_2$ mass of the galaxy excluding the nucleus, and $\mathfrak{M}$ the mass of the galaxy (including the nucleus).

We shall assume that the values of Oort's constants A and B at a distance R_0 from the center of the galaxy are known. This is equivalent to a knowledge of the force F_0 at a distance R_0, and its derivative F'_0.

We recall that expressions such as "distance of the sun from the center of the galaxy", "radius of the galaxy," "galaxy" etc., mean only that we are assuming numerical values for the parameters of the initial distribution which are equal to those accepted for the real galaxy.

The initial distribution. A spherically-symmetric, nonuniform system is fully defined when the density of matter $\rho(R)$ is given. Without assuming the form of $\rho(R)$ we can set up a system of equations which, using the given parameters of the initial distribution (r, R_0, F_0 and F'_0), will determine the remaining parameters. Let

$$\rho = \rho(R, \varkappa_1, \varkappa_2), \tag{3}$$

where κ_1 and κ_2 are the parameters to be determined.

We can then write down the following equations

$$\left.\begin{aligned} \rho(R_0, \varkappa_1, \varkappa_2) &= \rho_0; \\ \rho(r, \varkappa_1, \varkappa_2) &= \delta, \end{aligned}\right\} \tag{4}$$

where ρ_0 (density of matter at the distance R_0) and δ (constant density of the nucleus) are also unknown.

The force of attraction at the distance R_0 is given by

$$F_0 = \frac{G\,\mathfrak{M}_1}{R_0^2} + \frac{GM}{R_0^2}, \tag{5}$$

where

$$\mathfrak{M}_1 = \frac{4}{3}\pi r^3 \delta \tag{6}$$

$$M = 4\pi \int_r^{R_0} R^2 \rho(R, \varkappa_1, \varkappa_2)\, dR. \tag{7}$$

Taking into account (6) and (7) we can find expressions for F_0 and F'_0 in terms of the required parameters

$$\left.\begin{aligned} F_0 &= \frac{4}{3}\pi G \frac{r^3}{R_0^2}\delta + \frac{4\pi G}{R_0^2}\int_r^{R_0} R^2 \rho(R, \varkappa_1, \varkappa_2)\, dR. \\ F'_0 &= -\frac{8}{3}\pi G \left(\frac{r}{R_0}\right)^3 \delta + 4\pi G \rho_0 - \frac{8\pi G}{R_0^3}\int_r^{R_0} R^2 \rho(R, \varkappa_1, \varkappa_2)\, dR. \end{aligned}\right\} \tag{8}$$

The set of Equations (4) and (8) are a set of four equations involving four unknowns (ρ_0, δ, κ_1 and κ_2) which with a given form of the function $\rho(R, \kappa_1, \kappa_2)$ can in principle be solved without difficulty.

The interpretation of observational data given by B. V. Kukarkin and P. P. Parenagoshows that the exponential distribution of density is in good agreement with observation (cf., for example, [2] Section 50). We shall therefore take

$$\rho(R, \varkappa_1, \varkappa_2) = \varkappa_1 e^{-\varkappa_2 R}. \tag{9}$$

Let r = 1.5 kparsec; R = 7.2 kparsec; R_G = 13 kparsec;
A = 19.5 km/sec/kparsec; B = − 13 km/sec/parsec.

On solving (4) and (8) we find

$$\mathfrak{M}_1 = 0.61 \cdot 10^{10} \text{ sun masses}$$
$$\mathfrak{M} = 10.7 \cdot 10^{10} \text{ sun masses}$$

i. e., the mass of the nucleus of the initial distribution is ~ 5.7% of the whole mass of the distribution.

The nonuniform model. We shall now consider what will be the effect on the relative mass of the nucleus $\mathfrak{M}_1/\mathfrak{M}$ of the adoption of a uniform model for the nonuniform initial distribution under consideration. In this case instead of (6) and (7) we have

$$\mathfrak{M} = \frac{4}{3}\pi r^3 \delta; \tag{10}$$

$$M = \frac{4}{3}\pi (R_0^3 - r^3)\bar{\rho}, \tag{11}$$

where $\bar{\rho}$ is the density of the uniform mass outside the nucleus.

Equation (8) now takes the form

$$\begin{aligned} F_0 &= \frac{4}{3}\pi G \frac{r^3}{R_0^2}\delta + \frac{4}{3}\pi G\left(R_0 - \frac{r^3}{R_0^2}\right)\bar{\rho}, \\ F_0' &= -\frac{8}{3}\pi G\left(\frac{r}{R_0}\right)^3 \delta + \frac{4}{3}\pi G\left(1 + \frac{2r^3}{R_0^3}\right)\bar{\rho}. \end{aligned} \tag{12}$$

Solving Equation (12) with respect to δ and $\bar{\rho}$ we find finally the following values for the mass of the nucleus $\mathfrak{M}_1$ and the total mass of the model:

$$\mathfrak{M}_1 = 0.72 \cdot 10^{11} \text{ sun masses}$$
$$\mathfrak{M} = 1.79 \cdot 10^{11} \text{ sun masses}$$

i. e., the mass of the nucleus is 40.5% of the total mass.

Comparing the magnitude of the relative mass of the nucleus $\mathfrak{M}_1/\mathfrak{M}$ for the initial distribution with that calculated for an initial distribution using a uniform model, we see that the adoption of a uniform model for a spherically symmetrical nonuniform distribution leads to an overestimation of the relative nuclear mass by a factor greater than seven.

It is easy to see using Equation (8) that in the case of a spherically symmetrical distribution of density we have

$$\rho(R) = \frac{2F(R) + R\,\dfrac{dF(R)}{dR}}{4\pi G R}\,. \qquad (13)$$

The latter may be transformed into the following form:

$$\rho(R) = \frac{(B-A)(A+3B)}{4\pi G}\,. \qquad (13a)$$

(Equation (13a) may be convenient in structural studies and the dynamics of globular clusters and galaxies of type EO.)

Using Equation (13a) with $R = R_0$ we have

$$\rho(R_0) = \rho_0 = 7.96\cdot 10^{-25}\ \mathrm{g/cm^3}$$

This value of ρ_0 is smaller by a factor of 7-8 than the observed density in the vicinity of the sun: $\rho_0 = 5.5 - 6.6\cdot 10^{-24}\ \mathrm{g/cm^3}$ [5].

This discrepancy shows that the density distribution in a real galaxy is far from spherical.

The quantity κ_2 in (9) is, clearly, $-\dfrac{d\ln\rho}{dR}$. In the initial distribution it is equal to 0.63 kparsec^{-1}, which leads to the following value for the logarithmic gradient of density: $-\dfrac{d\log\rho}{dR} = 0.27$ kparsec^{-1}.

Observations give the following values for $-\dfrac{d\log\rho}{dR}$ ([2] Section 50):

0.25-0.27 kparsec^{-1} (spherical component),
0.11-0.16 kparsec^{-1} (plane component).

The agreement between the numerical value of the logarithmic density gradient in the galactic plane and that observed for stars of the spherical component leads us to the conclusion that the dynamic effect of the spherical component is far greater than is usually supposed.

3. Let the initial distribution be a nonuniform spheroid. Let us find the magnitude of the relative mass of the nucleus $\mathfrak{M}_1/\mathfrak{M}$. We shall consider a nonuniform spheroid with an exponential density distribution in the equatorial plane in accordance with Equation (9). If one assumes that the surfaces of equal density are similar spheroids then it is easy to find the value of the density at any point in the body. The nucleus will be assumed to be uniform. Let r = 1.5 kparsec; R_0 = 8 kparsec; A = 19.5 km/sec/kparsec; B = − 13 km/sec/kparsec.

The following notation is employed: $\underline{r}$ is the radius of the equatorial section through the nucleus; R is the radius vector in the equatorial plane of a spheroid; M is the mass contained within a spheroid with semimajor axis R_0; $\underline{e}$ is the eccentricity of the meridian section through the spheroid; ϵ is the compression ($\epsilon = \sqrt{1-e^2}$).

We emphasize once more that we are not attempting to model the galaxy, although the choice of initial data may suggest this. We are considering a fully defined initial distribution of matter with the object of discovering what errors are introduced by assuming a uniform model for it.

Undoubtedly, the initial spheroidal distribution under consideration is nearer to the structure of the real galaxy than the spherical distribution, although it does not appear to be in the best agreement with observational data on the galactic structure. For example, it does not take into account the observed compression of the nucleus [7], and it excludes the possibility of taking into account changes in compression in the galaxy.

We can get an idea of how near the given scheme is to a real galaxy from the numerical values of the calculated parameters (e. g., compression ϵ and the logarthmic density gradient in the galactic plane), which are sufficiently well known for the galaxy and are not used in the setting up of the model.

Initial distribution. We shall now proceed to determine the main parameters of the initial distribution: κ_1, κ_2, δ and ϵ. Having obtained these parameters we shall find the relative mass of the nucleus $\mathfrak{M}_1/\mathfrak{M}$. Using the calculated value of ϵ we shall obtain the uniform model of our initial distribution and the relative mass of the nucleus for this model, and hence we shall be able to find by direct comparison the errors introduced into the nuclear mass by the choice of the uniform model.

Since the compression ϵ is taken as unknown in the initial distribution, we shall assume that the density at a distance R_0 is known. We shall take it (using observational data on a real galaxy) to be $6.0 \cdot 10^{-24}$ g/cm^3.

We shall assume that r, R_0, ρ_0, F_0 and F'_0 are known. The system of equations from which we shall determine κ_1, κ_2, δ and ϵ may be written in the form

$$\left.\begin{aligned} \delta &= \varkappa_1 e^{-\varkappa_2 r}, \\ \rho_0 &= \varkappa_1 e^{-\varkappa_2 R_0}, \\ F_0 &= F_1(R_0) + F_2(R_0), \\ F'_0 &= F'_1(R_0) + F'_2(R_0), \end{aligned}\right\} \tag{14}$$

where $F_1(R)$ and $F_2(R)$ are the forces due to the nucleus and the spheroid, respectively.

The attraction due to the nonuniform spheroid, with the nucleus removed, can be expressed (assuming exponential distribution of density in the equatorial plane and spheroidal similarity of layers of equal density) in the form

$$F_2(R_0) = 4G\varkappa_1 \int\limits_{\varphi=0}^{\pi} \int\limits_{l=0}^{r} \int\limits_{z=\epsilon\sqrt{r^2-l^2}}^{\epsilon\sqrt{R_0^2-l^2}} \frac{(R_0 - l\cos\varphi)\, e^{-\varkappa_2\sqrt{l^2+z^2/\epsilon^2}}\, l\,dl\,d\varphi\,dz}{(z^2 + R_0^2 + l^2 - 2R_0 l\cos\varphi)^{3/2}} +$$

$$+ 4G\varkappa_1 \int\limits_{\varphi=0}^{\pi} \int\limits_{l=r}^{R_0} \int\limits_{z=0}^{\epsilon\sqrt{R_0^2-l^2}} \frac{(R_0 - l\cos\varphi)\, e^{-\varkappa_2\sqrt{l^2+z^2/\epsilon^2}}\, l\,dl\,d\varphi\,dz}{(z^2 + R_0^2 + l^2 - 2R_0 l\cos\varphi)^{3/2}}.$$

This integral cannot be evaluated completely by analytical methods and numerical methods must therefore be used. We shall evaluate this integral approximately by replacing it by a sum of integrals. Let us divide the spheroid into seven similar spheroidal layers (not counting the nucleus). We shall assume the density in each layer to be constant. We shall retain the overall exponential law of density distribution (Fig. 1).

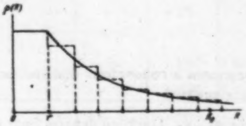

Fig. 1. The continuous line represents the true "integral" density distribution; the dotted line represents the density distribution assumed in the evaluation of the sum.

To simplify the computations we calculated the full attraction by summing up the attractions due to a series of superimposed uniform spheroids, instead of calculating the sum of attractions due to the separate layers. The density of each of the spheroids is equal to the height of a "step" on the density curve, i. e., the difference in the density of successive spheroidal layers. For this reason, the density of the nucleus δ, for example, is represented as a sum of densities of eight superimposed spheroids.

A uniform compressed spheroid attracts a point mass lying in its equatorial plane with a force given by

$$F(R) = 2\pi G\rho \frac{a\sqrt{1-e^2}}{e^2}\left[\frac{R}{ae}\arcsin\frac{ae}{R} - \sqrt{1-\frac{a^2e^2}{R^2}}\right], \tag{15}$$

where a is the semimajor axis of the equator section through the spheroid, e is the eccentricity of a meridian section through the spheroid, ρ is the density of the spheroid. The force is assumed positive in the direction towards the center of the spheroid.

Putting $\frac{ae}{R} = u$, we obtain

$$F(R) = 2\pi G\rho \frac{\sqrt{1-e^2}}{e^2}\frac{a}{u}\left[\arcsin u - u\sqrt{1-u^2}\right],$$

$$F'(R) = 2\pi G\rho \frac{\sqrt{1-e^2}}{e^3}\left[\arcsin u - \frac{u(1+u^2)}{\sqrt{1-u^2}}\right].$$

Summing up over all the superimposed spheroids we finally have:

$$\frac{F_0}{2\pi G\rho_0} = \frac{\sqrt{1-e^2}}{e^2}\sum_{i=1}^{8}(a_i-b_i)c_i, \quad . \quad \frac{F'_0R_0}{2\pi G\rho_0} = \frac{\sqrt{1-e^2}}{e^2}\sum_{i=1}^{8}(a_i-d_i)c_i, \tag{16}$$

where a_i, b_i, c_i and d_i are functions of κ_2, e and other known quantities.

The set of Equations (16) is solved by a graphiconumerical method in which the right-hand sides of (16) are tabulated for different κ_2 and e.

Using the values of κ_1, κ_2, δ and e we find the following values for the initial distribution:

$$\varepsilon = 1/7.1,$$
$$\mathfrak{M}_1 = 3.47\cdot 10^9 \text{ sun masses},$$
$$\mathfrak{M} = 1.26\cdot 10^{11} \text{ sun masses}.$$

Thus the mass of the nucleus $\mathfrak{M}_1$ is 2.75% of the total mass of the initial nonuniform spheroidal distribution.

Uniform model. If we represent the initial distribution by a uniform model, i. e., a uniform spheroid with the same compression and a nucleus, we obtain the following values for the masses (using the method discribed above):

$$\mathfrak{M}_1 = 0.70 \cdot 10^{11} \text{ sun masses},$$
$$\mathfrak{M} = 1.24 \cdot 10^{11} \text{ sun masses}.$$

Thus the mass of the nucleus is 56.5% of the total mass. An attempt to represent a nonunoform distribution by a uniform model leads to a twentyfold (!) overestimation of the mass of the nucleus.

We have shown above that in his account of the determination of the galaxy, Lindblad departs from the uniform spheroid by assuming $\epsilon \ll 1$ and actually using $\epsilon = 0$, and therefore goes over to the case of a uniform disc.

It is possible, without writing down long integrals, and without assuming $\epsilon \ll 1$, nor $\epsilon = 0$, to obtain a solution to the problem in its general form.

In fact, it is easy to show that (using the notation of [2])

$$F_{02} = \frac{2\pi G\rho_0 (1 + l^2)}{l^3}\left(\arctan l - \frac{l}{1 + l^2}\right), \tag{17}$$

where $l = \frac{\sqrt{1 - \epsilon^2}}{\epsilon}$ (cf. [16]).

Let

$$\varphi(l) = \frac{3}{2}\frac{(1 + l^2)^{3/2}}{l^3}\left(\arctan l - \frac{l}{1 + l^2}\right); \tag{18}$$

then

$$M = \frac{F_{02} R_0^2}{G\varphi(l)}. \tag{19}$$

Substituting $F_{02} = 0.20 F_0$, which was obtained by P. P. Parenago [2] for a model of a uniform spheroid with a nucleus, we have

$$M = \frac{3.62 \cdot 10^{43}}{\varphi(l)}. \tag{20}$$

The total mass of the spheroid can be found from the expression

$$\mathfrak{M} = \mathfrak{M}_1 + M\left(\frac{R_G}{R_0}\right)^3 = \mathfrak{M}_1 + 5.93\, M. \tag{21}$$

Let us construct a table of $M(\epsilon)$ and $\mathfrak{M}_2(\epsilon)$.

The case considered by P. P. Parenago, i. e., the case of a plane disc, is given in the last line of Table 1. In the latter case we have $\mathfrak{M}_1/\mathfrak{M} = 0.60$, i. e., the mass of the nucleus $\mathfrak{M}_1$ is 60 % of the total mass of the galaxy. We note that the effect of the compression ϵ on the mass of the spheroid is relatively small.

Summing up the above results we conclude that attempts to study nonuniform distributions of mass in the galaxy using a uniform model lead to manifold overestimation of the relative mass of the nucleus of the galaxy.

TABLE 1

ε	l	$\varphi(l)$	M	$\mathfrak{M}_2$	Remarks
1	0	1.000	$3.62 \cdot 10^{43}\,\mathrm{g} = 1.81 \cdot 10^{10}\, m_{\odot}$	$10.7 \cdot 10^{10}\, m_{\odot}$	φ_0 is found from L'Hospital's rule
$^1/_2$	$\sqrt{3}$	1.418	2.55 1.28	7.6	
$^1/_3$	$2\sqrt{2}$	1.641	2.21 1.10	6.5	
$^1/_4$	$\sqrt{15}$	1.778	2.04 1.02	6.0	
$^1/_8$	$\sqrt{63}$	2.029	1.78 0.89	5.3	
0	∞	2.356	$1.54 \cdot 10^{43}\,\mathrm{g} = 0.77 \cdot 10^{10}\, m_{\odot}$	$3.6 \cdot 10^{10}\, m_{\odot}$	$\varphi(\infty) = 3/4$ is found from L'Hospital's rule

4. The method which was used in studying the model of a nonuniform spheroid with constant compression can be used to construct a model which will approximate to the real galaxy.

The galaxy is looked upon as a system consisting of a uniform spheroidal nucleus and a variable density spheroid. This spheroid is conditionally divided into two spheroidal parts, one inside the other, with a dividing spheroidal layer passing through the sun (R_0 = 8 kparsec) between them. The compression ϵ is assumed to be variable in the inner spheroid and constant in the outer one. This assumption means that we need not consider the gravitational effect of masses which lie beyond the limits of the spheroidal layer through the sun. Thus $\epsilon = \epsilon\,(R)$ (R is the radius vector in the equatorial plane) where

$$\begin{aligned} \varepsilon &= \varepsilon_r = \text{const} \quad \text{for} \quad R \leqslant r, \\ \varepsilon &= aR + b \qquad \text{for} \quad r \leqslant R \leqslant R_0, \\ \varepsilon &= \varepsilon_0 = \text{const} \quad \text{for} \quad R \geqslant R_0. \end{aligned}$$

The quantities a and b are found from boundary conditions to be equal to

$$a = \frac{\varepsilon_0 - \varepsilon_r}{R_0 - r}\,, \quad b = \frac{\varepsilon_r R_0 - \varepsilon_0 r}{R_0 - r}\,.$$

where ϵ_r is assumed equal to 1/2 (cf. [4] or [7]).

Such a model of the galaxy allows us to limit ourselves to a consideration of the densities in a "cut" through the galaxy along the equatorial plane. Suppose that the density at some point (x, y, z) in the galaxy is given. It is clear that the surfaces of equal density in our model are surfaces of revolution. We have to find that value of R which is equal to the radius of an equatorial section through a surface of equal density passing through the point (x, y, z):

$$\frac{x^2 + y^2}{R^2} + \frac{z^2}{R^2\,(ab + b)^2} = 1. \tag{22}$$

Equation (22) is of the fourth degree in R. Using the theorem of Descartes we know that there is at least one positive root of this equation and we shall denote it by u. It is then possible to write

$$\rho\,(x, y, z) = \varkappa_1 e^{-\varkappa_2 u}, \tag{23}$$

which can be used to determine the density at any point.

We recall that the previous paragraphs of this paper were mainly devoted to the problem of when it is allowable to represent a nonuniform distribution of matter by a uniform model. It we take the initial characteristics of a nonuniform distribution to be equal to those obtained from observations we can obtain as an additional result an estimate of how closely the nonuniform distribution under consideration approaches the real galaxy.

The aim of the present paragraph is to discuss our ideas on the structure of the galaxy and as an additional result obtain an estimate of the accuracy obtained when this model is approximated by a uniform spheroid.

Thus, the quantities r, R_0, ρ_0, ϵ_r, F_0 and F'_0 are considered to be known. The quantities κ_1, κ_2, δ and ϵ_0 are determined.

Let us construct a table of the quantities $M(\epsilon)$ and $\mathfrak{M}_2(\epsilon)$.

Using a method analogous to that of the previous paragraph we have

$$\frac{F_0}{2\pi G R_0 \rho_0} = \sum_{i=1}^{8} \gamma_i (a_i - b_i) c_i, \quad \frac{F'_0}{2\pi G \rho_0} = \sum_{i=1}^{8} \gamma_i (a_i - d_i) c_i, \tag{24}$$

where a_i, b_i, c_i, d_i, and γ_i are functions of κ_2, ϵ_0 and known quantities. We find

$$\varkappa_2 = 0.36 \text{kps}^{-1} \quad \text{i. e.,} \; -\frac{d \log \rho}{dR} = 0.16 \text{ kps}^{-1}$$

$\epsilon_0 = 1/9$, hence $\epsilon(R) = -0.0598 R + 0.5897$ where $r \le R \le R_0$);

$$\varkappa_1 = 1.07 \cdot 10^{-22} \text{g/cm}^3; \quad \delta = 0.62 \cdot 10^{-22} \text{ g/cm}^3.$$

Finally we find masses of the nucleus and the galaxy to be

$$\mathfrak{M}_1 = 0.64 \cdot 10^{10} \text{ sun masses,}$$
$$\mathfrak{M} = 1.82 \cdot 10^{11} \text{ sun masses,}$$

i. e., the mass of the nucleus is 3.5% of the total mass of the galaxy.

If we recall that the adoption of a uniform spheroidal model gave a nuclear mass which was 60% of the total mass we find that, by assuming a uniform model for the given distribution, the relative mass of the nucleus is overestimated by a factor of 17.

5. No model can pretend to cover the entire range of observed facts. For this reason any approximation to the real galaxy must involve the assumption of a finite number of parameters, best known from observations. The number of the given parameters must be such that the model is fully defined by them. The remaining parameters of the model are then uniquely determined. These parameters give new information on the structure of the galaxy which could not be deduced with confidence directly from observations.

Table 2 summarizes the main results of the present work.

Column (VII) in Table 2 shows that the uniform model cannot be used, even in the first approximation, for studying the distribution of mass and, consequently, the dynamic characteristics of the galaxy.

TABLE 2

No.	Model	Log. density gradient $\frac{d\log\rho}{dR}$ (kparsec^{-1})	Mass of nucleus $\mathfrak{M}_1$ (sun masses)	Total mass of galaxy $\mathfrak{M}$ (sun masses)	Relative mass of nucleus $\alpha = (\mathfrak{M}_1/\mathfrak{M})$	Overestimation of the relative mass of the nucleus by assuming a uniform model (α_{unif}/α)	Compression ϵ
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
1	Nonuniform spheres with spherical uniform nucleus	-0.27	$0.61 \cdot 10^{10}$	$1.07 \cdot 10^{11}$	5.7%	7 times	1
2	Nonuniform spheroid with a spheroidal uniform nucleus (constant compression)	-0.20	$0.35 \cdot 10^{10}$	$1.26 \cdot 10^{11}$	2.8%	20.5 times	1/7.1
3	Nonuniform spheroid with a spheroidal uniform nucleus (variable compression)	-0.16	$0.64 \cdot 10^{10}$	$1.82 \cdot 10^{11}$	3.5%	17 times	$\epsilon(R) = aR + b$ for $r \leq R \leq R_0$ $\epsilon(r) = 1/2$ $\epsilon(R_0) = 1/9$

Model No. 3, in which the initial parameters represent most of the observed characteristics of the galaxy, allows us to estimate the masses of the galaxy and its nucleus, the fall-off in the density in the plane of the galaxy and changes in the concentration of matter in the galactic plane as a function of distance from the galactic center.

Shternberg State Astronomical Institute

Received July 9, 1956

LITERATURE CITED

[1] B. Lindblad, Handbuch der Astrophysik 5, 937 (1933).

[2] P. P. Parenago, A Course of Stellar Astronomy,* 3rd edition (State Theor. Lit. Press, Moscow, 1954).

[3] K. F. Bottlinger, Veröffentl. Berlin-Babelsberg 10, 2 (1933).

[4] P. P. Parenago, Astron. J. (USSR) 25, 306 (1948).

---

* In Russian.

[5] P. P. Parenago, Astron. J. (USSR) 29, 245 (1952).

[6] L. N. Sretenskii, Theory of Newtonian Potential * (State Tech. and Theor. Lit. Press, Moscow-Leningrad, 1946).

[7] P. N. Kholopov, Astron. J. (USSR) 27, 110 (1950).

---

* In Russian.

PERMANENT CONFIGURATIONS IN THE PROBLEM OF FOUR BODIES AND THEIR STABILITY

V. A. Brumberg

The stability of permanent configurations in the problem of four bodies is investigated for some particular cases. Section 1 contains an analysis of particular solutions in the problem of four bodies based on the work of Andoyer, Meyer, MacMillan and Bartky, Subbotin and Moulton. A solution of the problem of stability of circular orbits in the problem of N bodies is given in Section 2. Section 3 deals with the stability of particular solutions of the four-body problem in some special cases. Finally, an attempt is made to apply the theory of permanent configurations to the Trapezium of Orion.

We know that in the problem of three mass points which interact according to Newton's law there exist only two premanent configurations, the straight line and the equilateral triangle. In circular motion the first of these is unstable (in Liapunov's sense) for all mass values but can possess some conditional stability. The necessary condition for the stability of the second configuration is the inequality

$$\frac{m_1m_2+m_2m_3+m_3m_1}{(m_1+m_2+m_3)^2}<\frac{1}{27}. \qquad (1)$$

The stability of permanent configurations in the problem of four bodies is of interest and is the subject of the present paper. Despite the simple character of the investigation in principle, it could be carried through for only a few particular cases.

In Section 1 Lagrange's solutions in the problem of four points are analyzed on the basis of papers by Andoyer [1], Meyer [2], MacMillan and Bartky [3, 4], Subbotin [5] and Moulton [6]. Section 2 discusses principally Andoyer's results [7] concerning the stability of circular orbits in the problem of N bodies. In Section 3 stability in some very special cases is investigated. Finally, an unsuccessful attempt is made to apply the theory of permanent configurations to the Trapezium of Orion.

1. Permanent Configurations in the Problem of Four Bodies

The equations of motion of the problem of four bodies in a barycentric system of rectangular coordinates with the rotation components p, q, r along the x, y, z axes is as follows:

$$\left.\begin{aligned}
&\ddot{x}_i - 2r\dot{y}_i + 2\,q\dot{z}_i - (q^2 + r^2)\,x_i + (pq - \dot{r})\,y_i + (pr + \dot{q})\,z_i = \\
&\qquad = \sum_j{}' m_j \frac{x_j - x_i}{r_{ij}^3}, \\
&\ddot{y}_i - 2p\dot{z}_i + 2r\dot{x}_i - (r^2 + p^2)\,y_i + (qr - \dot{p})\,z_i + (qp + \dot{r})\,x_i = \\
&\qquad = \sum_j{}' m_j \frac{y_j - y_i}{r_{ij}^3}, \\
&\ddot{z}_i - 2q\dot{x}_j + 2p\dot{y}_i - (p^2 + q^2)\,z_i + (rp - \dot{q})\,x_i + (rq + \dot{p})\,y_i = \\
&\qquad = \sum_j{}' m_j \frac{z_j - z_i}{r_{ij}^3},
\end{aligned}\right\} \tag{1.1}$$

where $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2, \quad i = 1, \ldots, 4; \quad j = 1, \ldots, 4; \quad i \neq j$ (the gravitational constant is taken as unity).

A permanent configuration is one in which the ratio of the distances between bodies is constant during the motion, i. e., the following condition is satisfied:

$$\frac{x_i}{x_i^0} = \frac{y_i}{y_i^0} = \frac{z_i}{z_i^0} = \rho\,(t), \quad x_i^0,\, y_i^0,\, z_i^0 = \text{const.} \tag{1.2}$$

Equations (1.1) reduce to the following:

$$\left.\begin{aligned}
&x_i^0\,[\ddot{\rho} - (q^2 + r^2)\,\rho] + y_i^0\,[-2r\dot{\rho} + (pq - \dot{r})\,\rho] + z_i^0\,[2q\dot{\rho} + (pr + \dot{q})\,\rho] = \\
&\qquad = \frac{1}{\rho^2} \sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}}, \\
&x_0^i\,[2r\dot{\rho} + (qp + \dot{r})\,\rho] + y_i^0\,[\ddot{\rho} - (r^2 + p^2)\rho] + z_i^0\,[-2p\dot{\rho} + (qr - \dot{p})\,\rho] = \\
&\qquad = \frac{1}{\rho^2} \sum_j{}' m_j \frac{y_j^0 - y_i^0}{r_{ij}^{03}}, \\
&x_i^0\,[-2q\dot{\rho} + (rp - \dot{q})\,\rho] + y_i^0\,[2p\dot{\rho} + (rq + \dot{p})\,\rho] + z_i^0\,[\ddot{\rho} - (p^2 + q^2)\,\rho] = \\
&\qquad = \frac{1}{\rho^2} \sum_j{}' m_j \frac{z_j^0 - z_i^0}{r_{ij}^{03}}.
\end{aligned}\right\} \tag{1.3}$$

There are three possible cases:

I. All points of the system lie on a single straight line. For this line we take the x-axis and choose the y-axis so that q ≡ 0. From (1.3) we obtain

$$\begin{gathered} r\rho^2 = c = \text{const}, \\ \ddot{\rho} = -\frac{n^2}{\rho^2} + \frac{c^2}{\rho^3} \quad (p = 0), \end{gathered} \tag{1.4}$$

and the necessary condition for the existence of a straight-line configuration is

$$\sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}} = -n^2 x_i^0, \quad n = \text{const}. \tag{1.5}$$

Thus, when $c \neq 0$ all points describe similar Keplerian orbits (with equal eccentricities) whose focus is at the center of gravity of the system, and the points simultaneously occupy similar positions. Equations (1.5) for the problem of N bodies were first investigated by Moulton, who showed that there are $\frac{1}{2}$ N ! solutions.

II. The points of the system do not lie in a single plane. Since each of the coefficients of x^0, y^0, z^0 in (1.3) must have the form const/ρ^2 we easily obtain

$$\frac{p}{\alpha} = \frac{q}{\beta} = \frac{r}{\gamma} = \rho^{-3/2}, \quad \alpha, \beta, \gamma = \text{const},$$

where it follows that revolution occurs around a fixed axis; taking this as the z-axis we obtain

$$x_i^0 [\rho\ddot{\rho} - \rho^2 r^2] - y_i^0 \frac{d}{dt}(\rho^2 r) = \frac{1}{\rho} \sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}},$$

$$x_i^0 \frac{d}{dt}(\rho^2 r) + y_i^0 [\rho\ddot{\rho} - \rho^2 r^2] = \frac{1}{\rho} \sum_j{}' m_j \frac{y_j^0 - y_i^0}{r_{ij}^{03}},$$

$$z_i^0 \rho\ddot{\rho} = \frac{1}{\rho} \sum_j{}' m_j \frac{z_j^0 - z_i^0}{r_{ij}^{03}}.$$

Consequently,

$$\frac{d}{dt}(\rho^2 r) \sum_i m_i (x_i^{02} + y_i^{02}) = 0,$$

i. e., $\rho^2 r$ = const and $\rho^3 r^2$ = const. A case with ρ = const, r = const is impossible because we would then have

$$\sum_j{}' m_j \frac{z_j^0 - z_i^0}{r_{ij}^{03}} = 0.$$

Thus r = 0, $\ddot{\rho} = -n^2/\rho^2$ (straight-line Keplerian motion) and the necessary conditions for the existence of a solution are

$$\left.\begin{aligned} \sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}} &= -n^2 x_i^0, \\ \sum_j{}' m_j \frac{y_j^0 - y_i^0}{r_{ij}^{03}} &= -n^2 y_i^0, \\ \sum_j{}' m_j \frac{z_j^0 - z_i^0}{r_{ij}^{03}} &= -n^2 z_i^0; \end{aligned}\right\}$$

or in vector form

$$\sum_j{}' m_j\left(\frac{1}{r_{ij}^{03}} - \frac{n^2}{M}\right)\mathbf{r}_{ij}^0 = 0, \quad M = \sum_i m_i.$$

Since the vectors $\mathbf{r}_{ij}^0$ are not coplanar it is necessary that

$$r_{12}^0 = r_{13}^0 = r_{14}^0 = r_{23}^0 = r_{24}^0 = r_{34}^0, \quad n^2 = \frac{M}{r_{ij}^{03}}.$$

Thus the only permanent spatial configuration is a regular tetrahedron with arbitrary masses for the points of the system, and motion occurs along straight lines passing through the center of gravity in the same direction for all points.

III. The points of the system lie in a single plane but not on a single straight line.

Taking this plane as the xy plane in (1.3), we have

$$p = \alpha\rho^{-3/2}, \quad q = \beta\rho^{-3/2}, \quad (\alpha^2 + \beta^2)\dot{\rho} = 0.$$

The case $\alpha^2 + \beta^2 \neq 0$ leads to the condition

$$\sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}} = 0,$$

which is impossible. Thus p = q = 0, and the equations of motion are again given by (1.4), i. e., all four points move in a fixed plane along similar conic sections. The necessary conditions for the existence of plane configurations are

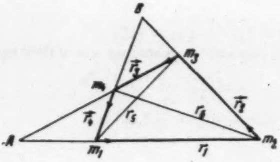

Fig. 1.

$$\left.\begin{aligned} \sum_j{}' m_j \frac{x_j^0 - x_i^0}{r_{ij}^{03}} &= -n^2 x_i^0, \\ \sum_j{}' m_j \frac{y_j^0 - y_i^0}{r_{ij}^{03}} &= -n^2 y_i^0. \end{aligned}\right\} \quad (1.6)$$

Let us consider, following MacMillan and Bartky, an arbitrary quadrilaterial with the sides, r_1, r_2, r_3, r_4. The directions of the corresponding vectors are shown in Fig. 1.

We denote by $\alpha_i r_i$ (i = 1, 2, 3, 4) the distance along $\mathbf{r}_i$ from the point m_i to the intersection with the opposite side (at point A or B); α_i is considered positive when the distance is measured in the same direction as $\mathbf{r}_i$; otherwise it is negative. It is easily established that

$$\mathbf{r}_{13}=\frac{1}{\alpha_2}[\beta_2\mathbf{r}_1+\beta_4\mathbf{r}_4],\quad \mathbf{r}_{31}=\frac{1}{\alpha_4}[\beta_4\mathbf{r}_3+\beta_2\mathbf{r}_2];$$

$$\mathbf{r}_{24}=\frac{1}{\alpha_3}[\beta_3\mathbf{r}_2+\beta_1\mathbf{r}_1],\quad \mathbf{r}_{42}=\frac{1}{\alpha_1}[\beta_1\mathbf{r}_4+\beta_3\mathbf{r}_3];\quad \beta_i=\alpha_i-1. \tag{1.7}$$

Introducing the new constant $R_0 = n^2/M$, using (1.7) and setting

$$R_0=\frac{1}{r_0^3},\quad R_i=\frac{1}{r_i^3},\quad S_i=R_i-R_0,\quad r_5=|\mathbf{r}_{13}|,\quad r_6=|\mathbf{r}_{24}|,\quad i=1,\ldots,6$$

(1.6) can be given the following vector form:

$$\left.\begin{aligned}
&(\alpha_2S_1m_2+\beta_2S_5m_3)\,\mathbf{r}_1+(\beta_4S_5m_3-\alpha_2S_4m_4)\,\mathbf{r}_4=0,\\
&(\alpha_3S_2m_3+\beta_3S_6m_4)\,\mathbf{r}_2+(\beta_1S_6m_4-\alpha_3S_1m_1)\,\mathbf{r}_1=0,\\
&(\alpha_4S_3m_4+\beta_4S_5m_1)\,\mathbf{r}_3+(\beta_2S_6m_1-\alpha_4S_2m_2)\,\mathbf{r}_2=0,\\
&(\alpha_1S_4m_1+\beta_1S_6m_2)\,\mathbf{r}_4+(\beta_3S_6m_2-\alpha_1S_3m_3)\,\mathbf{r}_3=0.
\end{aligned}\right\} \tag{1.8}$$

Noting that

$$\frac{\beta_{i+1}}{\alpha_{i-1}}=-\frac{\alpha_i}{\beta_i}, \tag{1.9}$$

and considering that the vectors $\mathbf{r}_i$ are noncollinear, we obtain two equivalent systems:

$$\left.\begin{aligned}
&S_1\alpha_2m_2+S_5\beta_2m_3=0,\quad &&S_6\alpha_2m_2+S_3\beta_2m_3=0,\\
&S_2\alpha_3m_3+S_6\beta_3m_4=0,\quad &&S_5\alpha_3m_3+S_4\beta_3m_4=0,\\
&S_3\alpha_4m_4+S_5\beta_4m_1=0,\quad &&S_6\alpha_4m_4+S_1\beta_4m_1=0,\\
&S_4\alpha_1m_1+S_6\beta_1m_2=0,\quad &&S_5\alpha_1m_1+S_2\beta_1m_2=0
\end{aligned}\right\} \tag{1.10}$$

from which it follows necessarily that

$$S_1S_3=S_2S_4=S_5S_6 \tag{1.11}$$

and

$$(S_1-S_2)(S_5-S_3)(S_4-S_6)=(S_1-S_6)(S_5-S_2)(S_4-S_3).$$

Eliminating R_0 from (1.10), we have the following system, in which the coefficients of the masses depend only on the mutual separations:

$$\left.\begin{aligned}
&(S_1-S_6)\,\alpha_2m_2+(S_5-S_3)\,\beta_2m_3=0,\\
&(S_2-S_5)\,\alpha_3m_3+(S_6-S_4)\,\beta_3m_4=0,\\
&(S_3-S_6)\,\alpha_4m_4+(S_5-S_1)\,\beta_4m_1=0,\\
&(S_4-S_5)\,\alpha_1m_1+(S_6-S_2)\,\beta_1m_2=0.
\end{aligned}\right\} \tag{1.12}$$

The determinant is

$$\Delta = \begin{vmatrix} 0 & (S_1 - S_6)\alpha_2 & (S_5 - S_3)\beta_2 & 0 \\ 0 & 0 & (S_2 - S_5)\alpha_3 & (S_6 - S_4)\beta_3 \\ (S_5 - S_1)\beta_4 & 0 & 0 & (S_3 - S_6)\alpha_4 \\ (S_4 - S_5)\alpha_1 & (S_6 - S_2)\beta_1 & 0 & 0 \end{vmatrix} =$$

$$= (S_5 - S_1)(S_5 - S_3)(S_6 - S_2)(S_6 - S_4)\beta_1\beta_2\beta_3\beta_4 -$$
$$- (S_2 - S_5)(S_4 - S_5)(S_1 - S_6)(S_3 - S_6)\alpha_1\alpha_2\alpha_3\alpha_4 = 0$$

in virtue of (1.9) and (1.11). It is easily seen that the necessary and sufficient conditions for the vanishing of all third order minors of Δ are

$$S_1 = S_2 = S_5, \quad S_3 = S_4 = S_6,$$

which corresponds to an equilateral triangle with three arbitrary but equal masses at its vertices and a fourth arbitrary mass m_4 at its center. For any other permissible quadrilateral the matrix of (1.12) will be of rank three, so that all masses will be expressed uniquely in terms of one of them.

We note in passing that according to Williams [8] even in permanent configurations of five bodies four masses can be expressed uniquely in terms of a fifth, except when four equal masses are located at the vertices of a square and the fifth, which is at the center,is arbitrary within a definite limit

$$0.439 < \frac{m_5}{m_1} \leqslant 1.414.$$

This is the essential difference of the present configurations from a permanent three-body configuration, where the masses of the system are arbitrary.

The analysis of the permissible quadrilaterals in the problem investigated by MacMillan and Bartky begins with the skew-symmetrical quadrilateral

$$r_1 = r_2, \quad r_3 = r_4 \quad (r_1 \geqslant r_3).$$

Here

$$m_2 = -\frac{\alpha_1}{\beta_1}\frac{R_5 - R_3}{R_1 - R_6}\,m_1, \quad m_3 = m_1, \quad m_4 = \frac{2 - \alpha_1}{1 - \alpha_1}\frac{R_5 - R_1}{R_6 - R_3}\,m_1,$$
$$r_5^2 = \frac{\alpha_1^2}{\alpha_1 - 1}\,r_1^2 - \frac{(2 - \alpha_1)^2}{\alpha_1 - 1}\,r_3^2, \quad r_6^2 = (1 - \alpha_1)(r_1^2 - r_3^2).$$

Introducing $\chi = r_3/r_1$ and noting that $-\frac{2\chi}{1-x} \leqslant \alpha_1 \leqslant \frac{2\gamma}{1+\chi} \leqslant 1,$ these authors obtain the following critical values (Fig. 2):

$$\left.\begin{aligned} &\text{(a)} \quad r_5 = r_3, \quad \chi = \frac{\pm\,\alpha_1}{\sqrt{3 - 3\alpha_1 + \alpha_1^2}}, \quad m_2 = 0; \\ &\text{(b)} \quad r_5 = r_1, \quad \chi = \frac{\sqrt{1 - \alpha_1 + \alpha_1^2}}{2 - \alpha_1}, \quad m_4 = 0; \end{aligned}\right\} \qquad (1.13)$$

(continued)

$$\left.\begin{array}{lll}
\text{(c)} & \alpha_1 = 0, & m_2 = 0; \\
\text{(d)} & r_6 = r_3, \quad \alpha_1 = \dfrac{1-2\chi^2}{1-\chi^2}, & m_4 = \infty; \\
\text{(e)} & r_6 = r_1, \quad \alpha_1 = -\dfrac{\chi^2}{1-\chi^2}, & m_2 = \infty.
\end{array}\right\} \quad \begin{array}{c}\textbf{(1.13)}\\ \textbf{(continued)}\end{array}$$

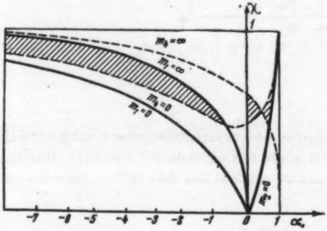

Fig. 2.

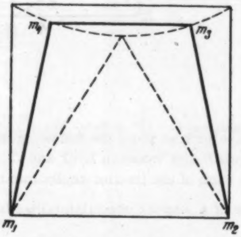

Fig. 3.

Each point of the shaded region gives a real solution of the problem. The intersection point of the two subregions where $\alpha_1 > 0$ has the coordinates $\alpha_1 = 1/2$, $\chi = 1/\sqrt{3}$ and corresponds to an equilateral triangle with its center at the point m_4. $\alpha_1 < 0$ corresponds to a convex quadrilateral and $\alpha_1 > 0$ corresponds to a concave quadrilateral (this means that when $\alpha_1 > 0$ the point m_4 is inside the triangle $m_1 m_2 m_3$). We note that $n^2 = MR_0$ can be written as

$$n^2 = 2m_1R_5 + m_2R_1 + m_4R_3. \tag{1.14}$$

Special cases of skew-symmetrical quadrilaterals are rhombuses and squares. In the first instance $m_1 = m_3$, $m_2 = m_4$; in the second instance $m_1 = m_2 = m_3 = m_4$.

The second class of permissible quadrilaterals with axial symmetry consists of equilateral trapezoids (Fig. 3). Here

$$r_2 = r_4, \; r_5 = r_6, \; m_1 = m_2, \; m_3 = m_4; \; \frac{m_3}{m_1} = \frac{r_1}{r_3}\sqrt{S_1/S_2}.$$

When an equilateral trapezoid changes from an equilateral triangle to a square the ratio m_3/m_1 changes continuously from 0 to 1. For all given values of m_1 and m_3 there exists one and only one equilateral trapezoid. In general all permissible quadrilaterals can be divided into those which are convex and those which are concave.

For any convex quadrilateral we have the inequalities

$$r_1, r_2, r_3, r_4 \leqslant r_0 \leqslant r_5, r_6,$$

with the two extreme sides lying opposite each other.

Let us consider Figure 4, where B_1OA_2, A_1OB_2 and $b_1a_1a_2b_2$ are semicircles of radius r_0, drawn with centers at the points m_1, m_2 and O. These determine the areas A and B. For each point $m_3 \subset B$ there exists one and only one point $m_4 \subset A$ (and no point outside A) such that the points m_1, m_2, m_3, m_4 form a permissible convex quadrilateral.

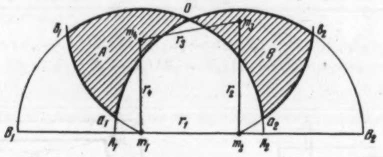

Fig. 4.

We now give without proof the following important properties of a permissible convex quadrilateral: the ratio of the diagonals lies between $1/\sqrt{3}$ and $\sqrt{3}$; each interior angle lies between 60° and 120°; finally, the diagonals divide each of the interior angles into two angles each of which is less than 60°.

In the case of a convex quadrilateral we have

$$r_1, r_2, r_5 \geqslant r_0 \geqslant r_3, r_4, r_6,$$

when only r_1 is a maximal side. In a construction similar to Figure 4 with some limitation of the regions (Fig. 5) the following theorem can be proved:

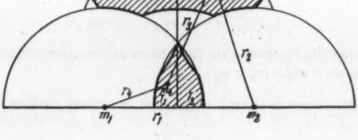

Fig. 5.

For each point $m_4 \subset T_1 + T_2$ there exists one and only one point $m_3 \subset S_1 + S_2$ such that the points m_1, m_2, m_3, m_4 form a permissible concave quadrilateral. When $m_3 \subset S_1$, we have $m_4 \subset T_1$.

MacMillan and Bartky prove, finally, that for any four masses in a given order there exists at least one permissible convex quadrilateral.

These are the principal results from the theory of permanent configurations in the problem of four bodies, which will be needed hereinafter.

2. The Stability of Circular Solutions in the Problem of Four Bodies

We shall first give the results which were obtained by Andoyer and Meyer and which apply to the problem of N bodies.

Everywhere in what follows we shall consider $\rho(t) = \text{const}$, i. e., only circular motions are considered. The plane of relative equilibrium of the given bodies will be the xy plane; the origin of coordinates will be the center of gravity of the system; the constant angular velocity of rotation of the x and y axes aroung the z axis will be n — the angular velocity of circular motion of the points. The equations of motion are

$$\left.\begin{aligned} \ddot{x}_s - 2n\dot{y}_s &= \frac{1}{m_s}\frac{\partial F}{\partial x_s}, \\ \ddot{y}_s + 2n\dot{x}_s &= \frac{1}{m_s}\frac{\partial F}{\partial y_s}, \\ \ddot{z}_s &= \frac{1}{m_s}\frac{\partial F}{\partial y_s}, \end{aligned}\right\} \qquad (2.1)$$

$$F = \frac{1}{2} n^2 \sum_s (x_s^2 + y_s^2) + \sum_{s,\,j} \frac{m_s m_j}{r_{sj}}.$$

In addition, $\sum\limits_s m_s x_s = 0,\ \sum\limits_s m_s y_s = 0,\ \ \sum\limits_s m_s z_s = 0$ и $\sum\limits_s \frac{m_s}{2}(\dot{x}_s^2 + \dot{y}_s^2 + \dot{z}_s^2) = F + \text{const}$ (Jacobian integral).

With the notation

$$\begin{aligned} u_s &= x_s + iy_s, \\ v_s &= x_s - iy_s. \end{aligned} \qquad i = \sqrt{-1},$$

we obtain

$$\left.\begin{aligned} \ddot{u}_s + 2ni\dot{u}_s &= \frac{2}{m_s}\frac{\partial F}{\partial v_s}, \\ \ddot{v}_s - 2ni\dot{v}_s &= \frac{2}{m_s}\frac{\partial F}{\partial u_s}, \\ \ddot{z}_s &= \frac{1}{m_s}\frac{\partial F}{\partial z_s}, \end{aligned}\right\} \qquad (2.2)$$

$$F = \frac{1}{2} n^2 \sum_s m_s u_s v_s + \sum_{sj} \frac{m_s m_j}{r_{sj}}; \qquad r_{sj}^2 = (u_s - u_j)(v_s - v_j) + (z_s - z_j)^2.$$

When $u_s = a_s$, $v_s = b_s$, $z_s = 0$ are the positions of relative equilibrium (when $a_s = x_s^0 + iy_s^0 = \text{const}$, $b_s = x_s^0 - iy_s^0 = \text{const}$ in the chosen coordinate system) we have

$$\left.\begin{aligned} \frac{\partial F}{\partial a_s} &= n^2 a_s - \sum_j m_j (a_s - a_j) a_{sj} = 0, \\ \frac{\partial F}{\partial b_s} &= n^2 b_s - \sum_j m_j (b_s - b_j) a_{sj} = 0, \end{aligned}\right\} \qquad (2.3)$$

$$a_{sj} = [(a_s - a_j)(b_s - b_j)]^{-3/2} = \frac{1}{r_{sj}^{\circ 3}}, \quad r_{sj}^{\circ 2} = (x_s^{\circ} - x_j^{\circ})^2 + (y_s^{\circ} - y_j^{\circ})^2.$$

Introducing into (2.2) $u_s = a_s + \xi_s$, $v_s = b_s + \eta_s$, $z_s = \zeta_s$, where ξ_s, η_s, ζ_s are assumed to be small, and neglecting terms of the second and higher orders, we obtain the following set of linear differential equations with constant coefficients:

$$\left.\begin{aligned} &\ddot{\xi}_s + 2ni\,\dot{\xi}_s = n^2\,\xi_s + \frac{1}{2}\sum_j \xi_j P_{sj} + \frac{3}{2}\sum_j \eta_j P'_{sj}\,, \\ &\ddot{\eta}_s - 2ni\dot{\eta}_s = n^2\,\eta_s + \frac{3}{2}\sum_j \xi_j P''_{sj} + \frac{1}{2}\sum_j \eta_j P_{sj}\,, \\ &\ddot{\zeta}_s + \sum_j \zeta_j P_{sj} = 0, \end{aligned}\right\} \qquad (2.4)$$

where

$$\begin{aligned} &P_{ss} = {\sum_j}' m_j a_{sj}, && P_{sj} = -\,m_j a_{sj}, && \sum_j P_{sj} = 0, \\ &P'_{ss} = {\sum_j}' m_j a_{sj}\,\frac{a_s - a_j}{b_s - b_j}\,, && P'_{si} = -\,m_j a_{sj}\,\frac{a_s - a_j}{b_s - b_j}\,, && \sum_j P'_{sj} = 0, \\ &P''_{ss} = {\sum_j}' m_j a_{sj}\,\frac{b_s - b_j}{a_s - a_j}\,, && P''_{sj} = -\,m_j a_{sj}\,\frac{b_s - b_j}{a_s - b_j}\,, && \sum_j P''_{si} = 0. \end{aligned}$$

(P' and P" are conjugate quantities).

The equilibrium conditions of (2.3) can be rewritten as

$$\left.\begin{aligned} n^2 a_s &= \sum_j a_j P_{sj} = \sum_j b_j P'_{sj}\,, \\ n^2 b_s &= \sum_j b_j P_{sj} = \sum_j a_j P''_{sj}. \end{aligned}\right\} \qquad (2.5)$$

I. The determination of ζ_s.

Let us set $\zeta_s = C_s e^{ht}$; then

$$C_s h^2 + \sum_j C_j P_{sj} = 0 \qquad (2.6)$$

and the secular equation will be

$$\Delta(h^2) = \begin{vmatrix} P_{11} + h^2 & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} + h^2 & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} + h^2 & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} + h^2 \end{vmatrix} = 0 \qquad (2.7)$$

Multiplying the rows by $\sqrt{m_1}$, $\sqrt{m_2}$, $\sqrt{m_3}$, $\sqrt{m_4}$ respectively, and dividing the columns by the same quantities, we obtain the determinant of a real symmetric matrix, from which it follows that all roots of (2.7) are real. In expanded form

$$\begin{gathered}
\Delta(h^2) = h^2[h^6 + H_1 h^4 + H_2 h^2 + H_3] = 0; \\
H_1 = m_1(a_{12} + a_{13} + a_{14}) + m_2(a_{21} + a_{23} + a_{24}) + m_3(a_{31} + a_{32} + a_{34}) + \\
+ m_4(a_{41} + a_{42} + a_{43}); \\
H_2 = m_1^2(a_{12}a_{13} + a_{13}a_{14} + a_{14}a_{12}) + m_2^2(a_{23}a_{24} + a_{24}a_{21} + a_{21}a_{23}) + \\
+ m_3^2(a_{34}a_{31} + a_{31}a_{32} + a_{32}a_{34}) + m_4^2(a_{14}a_{42} + a_{42}a_{43} + a_{43}a_{41}) + \\
+ m_1 m_2(a_{12}a_{23} + a_{23}a_{14} + a_{14}a_{12} + a_{12}a_{13} + a_{13}a_{24} + a_{24}a_{12}) + \\
+ m_1 m_3(a_{13}a_{32} + a_{32}a_{14} + a_{14}a_{13} + a_{13}a_{12} + a_{12}a_{34} + a_{34}a_{13}) + \\
+ m_1 m_4(a_{14}a_{43} + a_{43}a_{12} + a_{12}a_{14} + a_{14}a_{13} + a_{13}a_{42} + a_{42}a_{14}) + \\
+ m_2 m_3(a_{23}a_{31} + a_{31}a_{24} + a_{24}a_{23} + a_{23}a_{21} + a_{21}a_{34} + a_{34}a_{23}) + \\
+ m_2 m_4(a_{24}a_{41} + a_{41}a_{23} + a_{23}a_{24} + a_{24}a_{21} + a_{21}a_{43} + a_{43}a_{24}) + \\
+ m_3 m_4(a_{34}a_{41} + a_{41}a_{32} + a_{32}a_{34} + a_{34}a_{31} + a_{31}a_{42} + a_{42}a_{34}); \\
H_3 = M[m_1^2 a_{12}a_{13}a_{14} + m_2^2 a_{21}a_{23}a_{24} + m_3^2 a_{31}a_{32}a_{34} + m_4^2 a_{41}a_{42}a_{43} + \\
+ m_1 m_2 a_{12}(a_{14}a_{23} + a_{13}a_{24}) + m_1 m_3 a_{13}(a_{14}a_{32} + a_{12}a_{34}) + \\
+ m_1 m_4 a_{14}(a_{12}a_{43} + a_{13}a_{42}) + m_2 m_3 a_{23}(a_{24}a_{31} + a_{21}a_{34}) + \\
+ m_2 m_4 a_{24}(a_{21}a_{43} + a_{23}a_{41}) + m_3 m_4 a_{34}(a_{42}a_{31} + a_{41}a_{23})].
\end{gathered}$$

Thus the equation $\Delta(h^2) = 0$ has one zero and three negative roots. When $h^2 = 0$, the condition $\sum_s m_s C_s = 0$ can be satisfied only for $C_s = 0$, $s = 1, 2, 3, 4$, i. e., we again obtain Lagrange's solution. Multiplying the columns of (2.7) by a_1, a_2, a_3, a_4, respectively, and then by b_1, b_2, b_3, b_4, adding and taking (2.5) into account, we see that $h^2 = -n^2$ is a root of $\Delta(h^2) = 0$. It is a simple root if all points are on a single straight line (then all a_s are proportional to b_s); otherwise it is a double root. In the first case we can take $C_s = \mu a_s$, and in the second case $C_s = \mu a_s + \nu b_s$ (μ, ν = const).

Thus for any masses and initial perturbations perpendicular to the plane of unperturbed motion, the circular solutions are stable in first approximation and the perturbed motion with respect to the z-axis will be periodic.

II. The determination of ξ_s, η_s.

Setting $\xi_s = A_s e^{i\lambda t}$, $\eta_s = B_s e^{i\lambda t}$, we have

$$\begin{gathered}
A_s(\lambda + n)^2 + \frac{1}{2}\sum_j A_j P_{sj} + \frac{3}{2}\sum_j B_j P'_{sj} = 0, \\
B_s(\lambda - n)^2 + \frac{3}{2}\sum_j A_j P'_{sj} + \frac{1}{2}\sum_j B_j P_{sj} = 0
\end{gathered} \qquad (2.8)$$

and there is a characteristic equation (see page 69).

Therefore $\Delta(\lambda^2) = 0$ is an eighth degree equation in λ^2 with real coefficients. Adding the first and last four columns, we find that $\lambda^2 = n^2$ is a double root. Multiplying the first four columns by a_1, a_2, a_3, a_4 and the second set of four columns by b_1, b_2, b_3, b_4, adding and taking (2.5) into account, we obtain the following two columns:

$$\begin{array}{cc} a_s\left[\frac{1}{2}n^2+(\lambda+n)^2\right] & \frac{3}{2}n^2 a_s \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\ \frac{3}{2}n^2 b_s & b_s\left[\frac{1}{2}n^2+(\lambda-n)^2\right] \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \end{array} \qquad (s=1,\ 2,\ 3,\ 4)$$

which are proportional if

$$\left[\frac{1}{2}n^2+(\lambda+n)^2\right]\left[\frac{1}{2}n^2+(\lambda-n)^2\right]-\frac{9}{4}n^2=0, \text{ i. e., } \lambda^2=0,\ \lambda^2=n^2.$$

Thus, of the eight roots of (2.9) four are known: the triple root $\lambda^2 = n^2$ and the simple root $\lambda^2 = 0$.

Equation (2.9) can also be obtained in another form. Introducing into (2.1)

$$x_i = x_i^\circ + \varphi_i \qquad y_i = y_i^\circ + \psi_i \qquad z_i = z_i^\circ + \zeta$$

x_i^0 = const, y_i^0 = const, $z_i^0 = 0$ are Lagrange's solutions which satisfy (1.6), we obtain the system in the variations

$$\left.\begin{aligned} &\ddot{\varphi}_i - 2n\dot{\psi}_i = n^2\varphi_i + \sum_j \varphi_j P_{x^{ji}} + \sum_j \psi_j q_{ij}; \\ &\ddot{\psi}_i + 2n\dot{\varphi}_i = n^2\psi_i + \sum_j \varphi_j q_{ij} + \sum_j \psi_j P_{y^{ij}}; \\ &\ddot{\zeta}_i + \sum_j \zeta_j P_{ij} = 0; \end{aligned}\right\} \qquad (2.10)$$

where

$$P_{w^{ii}} = \sum_j{}' m_j a_{ij}\left[-1+3\frac{(w_j^0-w_j^0)^2}{r_{ij}^{02}}\right],\ P_{w^{ij}} = m_j a_{ij}\left[1-3\frac{(w_1^0-w_2^0)^2}{r_{ij}^{02}}\right],$$

$$\sum_j P_{w^{ij}} = 0, \quad w = x, y;$$

$$q_{ii} = 3\sum_j{}' m_j a_{ij}\frac{(x_j^0-x_j^0)(y_i^0-y_i^0)}{r_{ij}^{02}},\ q_{ij} = -3m_j a_{ij}\frac{(x_j^0-x_i^0)(y_j^0-y_i^0)}{r_{ij}^{02}},$$

$$\sum_j q_{ij} = 0,$$

Considering the equations for the determination of φ_i and ψ_i, we set $\varphi_i = A_i^\circ e^{gt}$, $\psi_i = B_i^\circ e^{gt}$. There is a characteristic equation (see page 69).

Equations (2.9) and (2.11) are two ways of writing the same equation. By setting $g^2 = -\lambda^2$, we can derive one from the other. The order of these determinants can easily be reduced by two through use of the double root $g^2 = -n^2$; but we did not succeed in reducing the order further from the knowledge of two additional roots $g^2 = -n^2$ and $g^2 = 0$. Generally speaking, a determinant of order 2N can be reduced to a determinant of order N by means of Schur's formula:

$$\Delta(\lambda^2) = \begin{vmatrix}
\frac{1}{2}P_{11}+(\lambda+n)^2 & \frac{1}{2}P_{12} & \frac{1}{2}P_{13} & \frac{1}{2}P_{14} & \frac{3}{2}P'_{11} & \frac{3}{2}P'_{12} & \frac{3}{2}P'_{13} & \frac{3}{2}P'_{14} \\
\frac{1}{2}P_{21} & \frac{1}{2}P_{22}+(\lambda+n)^2 & \frac{1}{2}P_{23} & \frac{1}{2}P_{24} & \frac{3}{2}P'_{21} & \frac{3}{2}P'_{22} & \frac{3}{2}P'_{23} & \frac{3}{2}P'_{24} \\
\frac{1}{2}P_{31} & \frac{1}{2}P_{32} & \frac{1}{2}P_{33}+(\lambda+n)^2 & \frac{1}{2}P_{34} & \frac{3}{2}P'_{31} & \frac{3}{2}P'_{32} & \frac{3}{2}P'_{33} & \frac{3}{2}P'_{34} \\
\frac{1}{2}P_{41} & \frac{1}{2}P_{42} & \frac{1}{2}P_{43} & \frac{1}{2}P_{44}+(\lambda+n)^2 & \frac{3}{2}P'_{41} & \frac{3}{2}P'_{42} & \frac{3}{2}P'_{43} & \frac{3}{2}P'_{44} \\
\frac{3}{2}P''_{11} & \frac{3}{2}P''_{12} & \frac{3}{2}P''_{13} & \frac{3}{2}P''_{14} & \frac{1}{2}P_{11}+(\lambda-n)^2 & \frac{1}{2}P_{12} & \frac{1}{2}P_{13} & \frac{1}{2}P_{14} \\
\frac{3}{2}P''_{21} & \frac{3}{2}P''_{22} & \frac{3}{2}P''_{23} & \frac{3}{2}P''_{24} & \frac{1}{2}P_{21} & \frac{1}{2}P_{22}+(\lambda-n)^2 & \frac{1}{2}P_{23} & \frac{1}{2}P_{24} \\
\frac{3}{2}P''_{31} & \frac{3}{2}P''_{32} & \frac{3}{2}P''_{33} & \frac{3}{2}P''_{34} & \frac{1}{2}P_{31} & \frac{1}{2}P_{32} & \frac{1}{2}P_{33}+(\lambda-n)^2 & \frac{1}{2}P_{34} \\
\frac{3}{2}P''_{41} & \frac{3}{2}P''_{42} & \frac{3}{2}P''_{43} & \frac{3}{2}P''_{44} & \frac{1}{2}P_{41} & \frac{1}{2}P_{42} & \frac{1}{2}P_{43} & \frac{1}{2}P_{44}+(\lambda-n)^2
\end{vmatrix} = 0 \tag{2.9}$$

$$\Delta(g^2) = \begin{vmatrix}
P_{x^{11}}+n^2-g^2 & P_{x^{12}} & P_{x^{13}} & P_{x^{14}} & q_{11}+2ng & q_{12} & q_{13} & q_{14} \\
P_{x^{21}} & P_{x^{22}}+n^2-g^2 & P_{x^{23}} & P_{x^{24}} & q_{21} & q_{22}+2ng & q_{23} & q_{24} \\
P_{x^{31}} & P_{x^{32}} & P_{x^{33}}+n^2-g^2 & P_{x^{34}} & q_{31} & q_{32} & q_{33}+2ng & q_{34} \\
P_{x^{41}} & P_{x^{42}} & P_{x^{43}} & P_{x^{44}}+n^2-g^2 & q_{41} & q_{42} & q_{43} & q_{44}+2ng \\
q_{11}-2ng & q_{12} & q_{13} & q_{14} & P_{y^{11}}+n^2-g^2 & P_{y^{12}} & P_{y^{13}} & P_{y^{14}} \\
q_{21} & q_{22}-2ng & q_{23} & q_{24} & P_{y^{21}} & P_{y^{22}}+n^2-g^2 & P_{y^{23}} & P_{y^{24}} \\
q_{31} & q_{32} & q_{33}-2ng & q_{34} & P_{y^{31}} & P_{y^{32}} & P_{y^{33}}+n^2-g^2 & P_{y^{34}} \\
q_{41} & q_{42} & q_{43} & q_{44}-2ng & P_{y^{41}} & P_{y^{42}} & P_{y^{43}} & P_{y^{44}}+n^2-g^2
\end{vmatrix} = 0 \tag{2.11}$$

$$\Delta = \begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} A & B \\ 0 & D - CA^{-1}B \end{vmatrix} = |A|\,|D - CA^{-1}B| = |AD - ACA^{-1}B|,$$

where A, B, C, D are quadratic matrices and $|A| \neq 0$. But in practice this method is ineffectual.

If the determinant (2.9) or (2.11) is calculated the characteristic equation of the system in the variations is

$$\Delta(g^2) = g^2 (g^2 + n^2)^3 (g^8 + ag^6 + bg^4 + cg^2 + d) = 0,$$

where $a = 5n^2 - (P_{11} + P_{22} + P_{33} + P_{44})$, but for the calculation of b, c, d one must know the sum of the principal minors of the second, third and fourth orders, respectively, of the determinant $\Delta(0)$.

It is known that the equation

$$a_0x^4 + 4a_1x^3 + 6a_2x^2 + 4a_3x + a_4 = 0 \tag{2.13}$$

then and only then has all real, negative and different roots when the following conditions are fulfilled:

$$a_0 > 0,\ a_1 > 0,\ a_2 > 0,\ a_3 > 0,\ a_4 > 0;$$
$$a_1^2 - a_0a_2 > 0;\ 12(a_1^2 - a_0a_2)^2 - a_0^2 f_1 > 0;\ D = f_1^3 - 27f_2^2 > 0,$$

where $f_1 = a_0a_4 - 4a_1a_3 + 3a_2^2$; $f_2 = a_0a_2a_4 + 2a_1a_2a_3 - a_0a_3^2 - a_1^2a_4 - a_2^3$. Setting $a_0 = 1$, $4a_1 = a$, $6a_2 = b$, $4a_3 = c$, $a_4 = d$, we obtain the necessary conditions for the stability of the position of relative equilibrium. When D = 0 Equation (2.13) has multiple roots.

We finally give the formulation of the problem of the stability of the libration points in a limited problem of four bodies. The equations of motion of a point of zero mass will be

$$\left.\begin{aligned} \ddot{u} + 2ni\,\dot{u} &= 2\frac{\partial F}{\partial v}, \\ \ddot{v} - 2ni\,\dot{v} &= 2\frac{\partial F}{\partial u}, \\ \ddot{z} &= \frac{\partial F}{\partial z}, \end{aligned}\right\}$$

$$F = \frac{1}{2}m^2uv + \sum_j{}' \frac{m_j}{r_{sj}}, \quad m_s = 0.$$

If $\xi = Ae^{i\lambda t}$, $\eta = Be^{i\lambda t}$, the characteristic equation will be

$$\lambda^2 - \lambda^2(2n^3 - P) + \left(\frac{1}{2}P + n^2\right)^2 - \frac{9}{4}P'P'' = 0, \tag{2.14}$$

$$P = P_{ss},\ P' = P'_{ss},\ P'' = P''_{ss}.$$

3. The Stability of Some Particular Circular Solutions of the Problem of Four Bodies

In the case of a straight-line configuration Equation (2.9) can be completely investigated. Taking $y_s^0 = 0$, $s = 1, 2, 3, 4$, we obtain $a_s = b_s = x_s^0$, $P_{sj} = P'_{sj} = P''_{sj}$. If h^2 is a root of (2.7), then according to (2.8)

$$m_s C_s h^2 + \sum_j m_j C_j P_{js} = 0 .$$

Multiplying the first four equations and then the second four equations of (2.8) by $m_s C_s$ and adding, we have

$$\left[(\lambda + n)^2 - \frac{1}{2} h^2\right] \sum_s m_s C_s A_s - \frac{3}{2} h^2 \sum_s m_s C_s B_s = 0 ;$$
$$\left[(\lambda - n)^2 - \frac{1}{2} h^2\right] \sum_s m_s C_s B_s - \frac{3}{2} h^2 \sum_s m_s C_s A_s = 0,$$

i. e., (2.9) in this case allows the roots of the equation

$$\lambda^4 - (h^2 + 2n^2)\lambda^2 + (n^2 + h^2)(n^2 - 2h^2) = 0. \tag{3.1}$$

If both roots of (3.1) are to be positive it is necessary that $-n^2 < h^2 < -\frac{8}{9} n^2$. But, as was proved by Meyer, all roots of (2.7) satisfy the condition $h^2 < -n^2$ (except, of course, $h^2 = 0$, $h^2 = -n^2$). Therefore (2.9) for a straight-line configuration has, in addition to $\lambda_{1,2} = 0$, $\lambda_{3,4,5} = n$, $\lambda_{6,7,8} = -n$, two pairs of real roots of opposite signs: $\lambda_9 = -\lambda_{10}$, $\lambda_{11} = -\lambda_{12}$, and four pure imaginary roots: $\lambda_{13} = i\alpha$, $\lambda_{14} = i\beta$, $\lambda_{15} = i\alpha$, $\lambda_{16} = -i\beta$. The existence of the last two roots proves that the straight-line configuration is unstable. Nevertheless, with suitable choice of the initial values the straight-line configuration possesses conditional stability, and there exist two families of periodic orbits close to the unperturbed positions of relative equilibrium.

It is also possible to calculate the coefficients of the characteristic equation in some very simple cases of plane configurations. The calculations are lengthy and will be omitted.

I. The square. With the notation

$$m = m_1 = m_2 = m_3 = m_4, \qquad a = a_{12} = a_{23} = a_{34} = a_{41},$$

we obtain

$$\Delta(g^2) = g^2 (g^2 + n^2)^3 \left[g^4 + \sqrt{2}\, mag^2 + \left(\frac{33}{2} + 4\sqrt{2}\right) m^2 a^2\right]\left[g^4 + n^2 g^2 + \right.$$
$$\left. + \frac{18}{\sqrt{2}} m^2 a^2 \right] = 0 \tag{3.2}$$

and

$$n^2 = ma\left(2 + \frac{1}{\sqrt{2}}\right).$$

Since the discriminants of the last two factors are

$$D_1 = -16(4+\sqrt{2})\,m^2a^2 < 0, \quad D_2 = \left(\frac{9}{2} - 34\sqrt{2}\right)m^2a^2 < 0,$$

each of these fourth-degree equations has roots of the forms $\alpha + i\beta,\ \alpha - i\beta,\ -\alpha - i\beta,\ -\alpha + i\beta$. Therefore (3.2) has four roots for g with positive real parts, from which the instability of this configuration follows. As in the case of the straight-line configuration, this instability is not of absolute character.

II. The rhombus. Considering that

$$m_1 = m_3, \qquad m_2 = m_4, \qquad n^2 = 2(m_1a_{12} + m_2a_{24}) = 2(m_1a_{13} + m_2a_{12}),$$

we obtain

$$\Delta(g^2) = g^2(g^2+n^2)^3\,[g^4 + g^2(2n^2 - M) + \\ + \left\{M^2\left(-2 + \frac{9}{16}r_{13}^2r_{24}^2\right) + Mn^2 + n^4\right\}\Big]\,[g^4 + n^2g^2 + A] = 0, \tag{3.3}$$

where

$$A = 36\sin^2\omega\cos^2\omega\,(m_1 - m_2)^2 + 36\,m_1m_2\,(a_{13}\cos^2\omega + a_{24}\sin^2\omega),$$

$$\omega = \arctan\frac{r_{13}}{r_{24}}, \quad 30^\circ \leqslant \omega \leqslant 60^\circ, \quad r_{12} = 1.$$

The discriminant of the first biquadratic equation is

$$D_1 = 9M^2\left(1 - \frac{1}{4}r_{13}^2r_{24}^2\right) - 8Mn^2 = M^2\left[8(1-R_0) + \left(1 - \frac{9}{4}r_{13}^2r_{24}^2\right)\right].$$

But from (1.11) we derive

$$R_0 = \frac{R_1R_3 - R_2R_4}{R_1 + R_3 - R_2 - R_4} = \frac{R_2R_4 - R_5R_6}{R_2 + R_4 - R_5 - R_6} = \frac{R_5R_6 - R_1R_3}{R_5 + R_6 - R_1 - R_3}. \tag{3.4}$$

Substituting in this case

$$1 - R_0 = \frac{1 - (r_{13}^3 + r_{24}^3) + r_{13}^3r_{24}^3}{2r_{13}^3r_{24}^3 - (r_{13}^3 + r_{24}^3)},$$

we find that $D_1 < 0$. Investigating the discriminant of the last factor in similar fashion we obtain $D_2 = n^4 - 4A < 0$. Thus in this case also the characteristic Equation (3.3) for g has four roots with positive real parts.

III. The equilateral triangle: three equal masses ($m_2 = m_3 = m_4$) at the vertices and a fourth mass (m_1) at the center.

We choose the units of length and mass so that

$$r_{23} = r_{34} = r_{42} = 1, \qquad n^2 = 3(m_2 + \sqrt{3}\,m_1)\,a_{23} = 1;$$

and let

$$M_1 = \frac{3\sqrt{3}}{2}\,m_1, \quad M_2 = (3\sqrt{3} - 1)\,m_2.$$

Then

$$\Delta(g^2) = g^2(g^2 + 1)^3\,[(g^4 + pg^2 + q)^2 + r^4 g^2] = 0 \tag{3.5}$$

or

$$\Delta(\lambda^2) = \lambda^2(\lambda^2 - 1)^3\,[\lambda^4 - p\lambda^2 + r^2\lambda + q]\,[\lambda^4 - p\lambda^2 - r^2\lambda + q] = 0, \tag{3.6}$$

where

$$p = 1 - \frac{3}{2}M_2, \quad r^2 = 3M_2, \quad q = \frac{9}{4}(1 + M_2) - \frac{9}{2}M_1(1 + 3M_2).$$

Let us consider the equation

$$f(x) = (x^2 + px + q)^2 + r^4 x = 0.$$

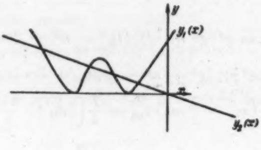

Fig. 6.

Analyzing the relation of the curve $y_1(x) = (x^2 + px + q)^2$ to the straight line $y_2(x) = -r^4 x$ (Fig. 6), we find that the necessary and sufficient conditions for the reality (and consequently the negative character) of the roots of $f(x) = 0$ are

$$\text{A)}\ p > 0; \quad \text{B)}\ q > 0; \quad \text{C)}\ \frac{p^2}{4} - q > \sqrt{\frac{p}{2}}\,r^2.$$

This means

$$\text{A) } M_2 < \frac{2}{3}; \quad \text{B) } M_1 < \frac{1+M_2}{2(1+3M_2)};$$

$$\text{C) } M_1 > \frac{3M_2\sqrt{2-3M_2} - \frac{9}{b}M_2^2 + 6M_2 + 4}{9(1+3M_2)}.$$

But conditions B) and C) lead to

$$27M_2^3 + 522M_3^2 + 36M_2 - 8 < 0,$$

from which $M_2 < 0.093860$. At the same time it follows from B) that $M_2 < \frac{6\sqrt{3}-5}{9} = 0.59915$.

Thus for any finite masses this configuration is unstable.

IV. Isosceles trapezoid. In this case $m_1 = m_2$, $m_3 = m_4$, $n^2 = (m_1 + m_3)(a_{12} + a_{13})$ (Fig. 3). The characteristic equation could not be investigated; we note only that in the limiting cases where the trapezoid degenerates into an equilateral triangle ($m_3 = 0$) or a square ($m_3 = m_1$) the configuration is unstable; in the first instance we have

$$\Delta(g^2) = g^2(g^4+1)^3\left(g^2+2g^2+\frac{27}{4}\right)^2 = 0; \quad (r_{12} = 1, \ m_1 = 1), \tag{3.7}$$

and the second case is discussed above.

V. Rhombohedron. The characteristic equation for this configuration could also not be investigated. We only note some critical cases corresponding to a), b) and c) of (1.13).

Let $r_{12} = r_{14}$, $r_{23} = r_{34}$, $m_2 = m_4$ (Fig. 1) and a point is found on the straight line m_2, m_4. Then with $m_3 = 0$ the configuration is unstable, since otherwise the collinear configuration of the three points m_2, m_1, m_4 would also be stable. This follows from the analysis of the characteristic equation:

$$\begin{gathered}\Delta(g^2) = g^2(g^2+n^2)^3\{[g^2 - n^2 - 2m_1a_{13} + \\ + 2m_2a_{23}(1-3\cos^2\beta)][g^2 - n^2 + m_1a_{13} + 2m_2a_{23}(1-3\sin^2\beta)] + \\ + 4n^2g^2\}\left[g^4 + g^2\left(m_1 - \frac{3}{2}m_2\right)a_{12} - \frac{7}{4}\left(3m_1 + \frac{17}{4}m_2\right)m_2a_{12}^2\right] = 0, \\ \beta = \arctan\frac{r_{12}}{r_{13}},\end{gathered} \tag{3.8}$$

the last factor of which has two pure imaginary roots for g and two roots which are equal in absolute magnitude but opposite in sign. The second biquadratic equation, which according to (2.14) determines the stability of the liberation point of m_3, was not investigated, because, according to Hüttenhain [9], the libration point in this case is unstable if $m_1/m_2 > 11.72\,138$ (libration points lying on the axis which passes through m_2 and m_4 are unstable).

Let us consider the equilateral triangle m_1, m_2, m_3 where $m_2 = m_3$ (Fig. 7).

The coordinate origin is taken at m_1. The fourth mass $m_4 = 0$ can be either outside or inside of this triangle.

Calculating (2.9), we have

$$\Delta(\lambda^2) = \lambda^2(\lambda^2 - n^2)^3 \Big\{\lambda^4 + \lambda^2(P_{44} - 2n^2) + \\ + \Big[\Big(n^2 + \frac{1}{2}P_{44}\Big)^2 - \frac{9}{4}P'_{44}P'_{44}\Big]\Big\}\Big[\lambda^4 - n^2\lambda^2 + \frac{27}{2}\Big(m_1 m_2 + \frac{m_2^2}{2}\Big)a_{12}^2\Big] = 0. \tag{3.9}$$

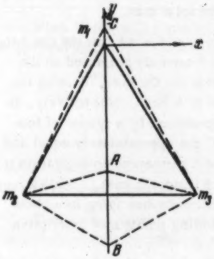

Fig. 7.

The problem of the stability of such configurations also divides into two problems: the stability of an equidistant configuration of three points where two of the three masses are equal, and the stability of the corresponding libration point in the limited problem of four bodies. The first follows from the fact that the positive discriminant of the last factor leads to the inequality

$$m_1^2 - 50m_1m_2 - 23m_2^2 > 0, \tag{3.10}$$

which also follows from (0.1)*when $m_2 = m_3$.

We denote the libration points by C, A and B. Each of the three systems of four mass points is stable if and only if, according to (3.10), $m_1/m_2 > 50.4558$, and the corresponding libration point is stable. The libration point on the axis $x = 0$ is determined by the following equation given by Meyer:

$$y + \frac{\sqrt{3}}{2}(1-\mu) - \frac{\mu}{y^2}\,sgn\,y - \frac{(1-\mu)\left(y + \frac{\sqrt{3}}{2}\right)}{(y^2 + \sqrt{3}\,y + 1)^{3/2}} = 0, \tag{3.11}$$

where $r_{12} = 1,\ m_2 = m_3 = \frac{1-\mu}{2},\ m_1 = \mu.$

Meyer also showed that there always exists a libration poing C between 0 and 1; the point B is between $-\sqrt{3}$ and -1, but when $\mu > 0.43242$ the libration point A ceases to exist. Hence it is clear that the configuration $m_1m_2Am_3$' is unstable, because in our case $\mu > 0.96187$.

It follows from (2.14) that for the stability of the libration point it is necessary that

$$2n^2 - P > 0,\ (P + 2n^2)^2 - 9P'^2 > 0,\ D = 9P'^2 - 8n^2P > 0. \tag{3.12}$$

In our case

$$n^2 = (m_1 + 2m_2)\,a_{12} = 1,\quad P = \frac{\mu}{|y|^3} + \frac{1-\mu}{(y^2 + \sqrt{3}y + 1)^{3/2}},$$

$$P' = -P + \frac{1-\mu}{2}\frac{1}{r_{12}^5},\quad r_{12}^2 = \frac{1}{4} + \left(y + \frac{\sqrt{3}}{2}\right)^2.$$

Giving a value to y, it is easy to find the corresponding values of μ and then the values of P and P'. Calculations showed that the configuration $m_1m_2Bm_3$ is stable for $m_1/m_2 \geq k = 367.0$ (with $y > -1.00153$), and that the configuration $m_1m_2Cm_3$ is unstable for any finite masses.

---

* As in original — Publisher's note.

4. Conclusion

We now make an attempt to apply the theory of permanent configurations of four bodies practically to the Trapezium of Orion, which consists of six main stars, two of them, E and F, differing considerably from the others A, B, C, D, by their large proper motions and radial velocities.

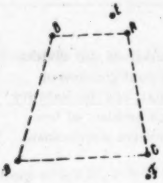

Fig. 8.

The projection of the Trapezium of Orion on the celestial sphere looks approximately like Figure 8 [10].

Observational data communicated by Professor P. P. Parenago is given in the table. The last measurements were made in 1920-1930 and the first measurements in 1840-1850; θ is the position angle, r_{ij} is a distance between stars in the plane of the figure; $\mathfrak{M}$ is the mass of a star in terms of the solar mass.

The distinct difference in the proper motions of stars E and F, on the one hand, and A, B, C, D on the other hand, suggests that E and F are only projected on the general background of the trapezoid. But even if this is not the case, by using the smallness of the masses of E and F and their closeness to A and C, respectively, the Trapezium of Orion can in first approximation be represented by a system of four bodies A, B,C and D, where the distances CD and AC are approximately equal and the masses D and A are also approximately equal. If this observed system formed a permanent configuration it would necessarily be a rhombohedron (or a quadrilateral very similar to it), which is not the case in reality. We now inquire whether the quadrilateral ABCD is only the projection of some rhombohedron lying in a plane that forms the angle i with the plane of the figure. Let us now consider the following problem of descriptive geometry: to construct a rhombohedron from its projection.

	AB	AC	AD	AE	BC	DB	DC	CF	$\mathfrak{M}$
r_{ij}	8″770 8″718 8″701 8″730	12″913 12″942 12″941 13″004	21″305 21″477 21″595 21″550	3″884 3″927 4″051 4″204	16″750 16″723 16″598 16″759	19″082 19″271 19″333 19″318	13″331 13″489 13″432 13″332	3″743 3″913 3″894 4″005	$A=10$ $B=6$ $C=21$
θ	32°30 32°28 32°34 32°20	131°36 131°28 131°55 131°20	95°55 95°57 95°73 95°58	352°28 352°02 351°38 350°90	163°29 163°66 163°91 163°97	299°59 299°43 299°69 300°00	240°58 240°97 241°13 241°36	122°10 122°70 122°59 121°60	$D=12$ $E=2.5$ $F=2.5$

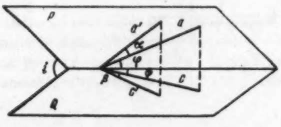

Fig. 9.

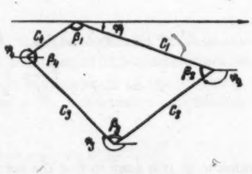

Fig. 10.

Let the projection of the segment a in plane P on plane Q be c (Fig. 9). In a parallel projection the segments and the angles between them are related by the formulas: [*]

$$\left.\begin{aligned} c = a\sqrt{1-\sin^2\varphi\sin^2 i}, \quad a = c\sqrt{1+\sin^2\Psi\,\mathrm{tg}^2\, i}, \quad \mathrm{tg}\,\Psi = \mathrm{tg}\,\varphi\cos i', \\ \sin\beta = \frac{\sin\alpha\cos i}{\sqrt{1-\sin^2\varphi\sin^2 i}\sqrt{1-\sin^2(\varphi+\alpha)\sin^2 i}}, \\ \sin\alpha = \frac{\sin\beta\sec i}{\sqrt{1+\sin^2\Psi\,\mathrm{tg}^2\, i}\sqrt{1+\sin^2(\Psi+\beta)\,\mathrm{tg}^2\, i}}. \end{aligned}\right\} \tag{4.1}$$

When in plane Q a quadrilateral is given with the sides c_1, c_2, c_3, c_4 and angles $\beta_1, \beta_2, \beta_3, \beta_4$ (Fig. 10), the position angles of its sides are related by

$$\Psi_n = 180° + \Psi_{n-1} - \beta_n = \Psi_1 + (n-1)\,180° - \sum_{k=2}^{n}\beta_k, \tag{4.2}$$

which is valid for any (convex or concave) n-sided figure.

Thus from a quadrilateral with the sides c_1, c_2, c_3, c_4 ($c_1 \neq c_2$, $c_3 \neq c_4$) we require the construction of a rhombohedron with the sides $a_1 = a_2$, $a_3 = a_4$, lying in a plane inclined at the angle i to the projection plane. The equalities $a_1 = a_2$, $a_3 = a_4$ lead to the equations

$$\begin{aligned} \mathrm{tg}^2\, i\,(c_1^2\sin^2\Psi_1 - c_2^2\sin^2\Psi_2) &= c_2^2 - c_1^2, \\ \mathrm{tg}^2\, i\,(c_3^2\sin^2\Psi_3 - c_4^2\sin^2\Psi_4) &= c_4^2 - c_3^2. \end{aligned} \tag{4.3}$$

The angle Ψ_1 is determined from

$$A\,\mathrm{tg}^2\Psi_1 + B\,\mathrm{tg}\,\Psi_1 + C = 0; \tag{4.4}$$

$$\begin{gathered} A = c_1^2\,(c_4^2 - c_3^2) - c_2^2\,(c_4^2 - c_3^2)\cos^2\beta_2 - c_3^2\,(c_2^2 - c_1^2)\cos^2(\beta_2+\beta_3) + \\ + c_4^2\,(c_2^2 - c_1^2)\cos^2\beta; \end{gathered}$$

$$B = c_2^2\,(c_4^2 - c_3^2)\sin 2\beta_2 + c_3^2\,(c_2^2 - c_1^2)\sin 2(\beta_2+\beta_3) + c_4^2\,(c_2^2 - c_1^2)\sin 2\beta_1;$$

$$C = -c_2^2\,(c_4^2 - c_3^2)\sin^2\beta_2 - c_3^2\,(c_2^2 - c_1^2)\sin^2(\beta_2+\beta_3) + c_4^2\,(c_2^2 - c_1^2)\sin^2\beta_1;$$

i. e.,

$$A + C = 0.$$

The solutions of (4.4) are

---

[*] In this paper 'tg' ≡ 'tan', 'ctg' ≡ 'cot' – Publisher.

$$\operatorname{tg} \Psi'_1 = \frac{-B + \sqrt{B^2 + 4A^2}}{2A}, \quad \operatorname{tg} \Psi''_1 = \frac{-B - \sqrt{B^2 + 4A^2}}{2A},$$

from which

$$\operatorname{tg} \Psi'_1 + \operatorname{ctg} \Psi''_1 = 0, \qquad \Psi''_1 - \Psi'_1 = \pm 90°.$$

Thus (4.4) has four roots $\Psi'_1, \Psi''_1, \Psi^*_1, \Psi^{**}_1$ separated from each other by 90°. Therefore

$$\sin^2 \Psi'_j = \cos^2 \Psi''_j, \; j = 1, \dots, 4$$

and from (4.3) we obtain

$$\operatorname{ctg}^2 i' + \operatorname{ctg}^2 i'' = -1. \tag{4.5}$$

$\operatorname{ctg}^2 i'' = 0$ gives the formal solution of the problem:

$$a_1 = a_2 = \infty, \quad a_3 = a_4 = \infty, \quad a_5 = 0, \quad a_6 = \infty$$

(a_5 and a_6 are diagonals).

Therefore the solutions of (4.5) are

$$\operatorname{ctg}^2 i'' = 0, \; \operatorname{ctg}^2 i' = -1. \tag{4.6}$$

When $i = i'$ we have

$$a_n = c_n \mid \cos \Psi'_n \mid, \quad n = 1, \dots, 4; \quad a_5 = 0, \quad a_6 = a_1 + a_3.$$

Consequently in this case we obtain a straight line as the limit of a rhombohedron with $a_1 = a_2$, $a_3 = a_4$ and the diagonal $a_5 \to 0$. Any quadrilateral ($c_1 \neq c_2$, $c_3 \neq c_4$) can be regarded as the projection of some segment, with the angle between the respective planes $i = i'$ determined from (4.6) and the position angle $\psi = \psi_1$ from (4.4).

The solutions (4.6), which are not rigorously derived from the theory, were checked a few times using numerical examples. One of these examples is the Trapezium of Orion. We first note that if we begin to construct the quadrilateral ABCD directly from observational data (using the sides r_{ij}) we will not obtain a closed figure because of errors. Therefore, taking definite values for CD and BC and the angles A, B, C, we determine all remaining elements of the quadrilateral. The final values are:

$$\begin{aligned}
AC &= c_1 = 12.886; & \angle A &= \beta_1 = 99°00'00''; \\
CD &= c_2 = 13.332; & \angle C &= \beta_2 = 110°09'36''; \\
DB &= c_3 = 19.237; & \angle D &= \beta_3 = 58°38'24''; \\
BA &= c_4 = 8.8878; & \angle B &= \beta_4 = 92°12'00''.
\end{aligned}$$

From these values we obtain $\psi_1' = 54°34'59''$, $\psi_1'' = -35°36'01''$ and the values of (4.6) subject to the accuracy of the calculations. We add that the straight-line configuration of three points which is obtained for $i = i'$:

$$a_1 = a_2 = 7.501, \quad a_3 = a_4 = 7.947; \quad a_5 = 0, \quad a_6 = 15.448$$

is not permanent for the given masses.

Thus within the framework of the present model (with the stars A and D of identical mass) the Trapezium of Orion cannot form a plane permanent configuration. It is only possible that the stars of the Trapezium form a spatial permanent configuration in the form of a regular tetrahedron, or that, according to the Pohlke-Schwarz theorem, any quadrilateral in a plane can be regarded as a parallel projection of any given tetrahedron.

In conclusion I wish to thank Professor G. N. Duboshin for directing this work.

P. K. Shternberg State Astronomical Institute

Received April 18, 1956

LITERATURE CITED

[1] Andoyer, Bull. Astron. 23, 50 (1906).

[2] Meyer, Ann. Observatoire Bordeaux 17 (1933).

[3] MacMillan, Dynamics of a Particle, (Russian Translation) (IIL, 1951) Ch. III.

[4] MacMillan and Bartky, Trans. Am. Math. Soc. 34, 838 (1932).

[5] M. F. Subbotin, Course in Celestial Mechanics Vol. II (ONTI, 1937) Ch. VI. *

[6] Moulton, Periodic Orbits (Washington, 1920) Ch. VIII.

[7] Andoyer, Bull. Astron. 23, 129 (1906).

[8] Williams, Trans. Am. Math. Soc. 44, 563 (1938).

[9] A. N. Huttenhain, 254, 281, 1935.

[10] P. P. Parenago, "An investigation of stars in the region of the nebula of Orion" [in Russian], Trudy P. K. Shternberg State Institute of Astronomy 25 (1954).

---

* In Russian.

ON SOME PECULIARITIES IN THE MOTIONS OF COMETS

P. G. Dukhnovskii

The changes in cometary orbits, caused by planetary perturbations, are considered. This problem is solved approximately by the doubly averaged limited circular problem of three points.

A formula for calculating the period of rotation of the line of apsides and formulas for the amplitude of the variation of inclination have been derived.

Calculations according to these formulas, in the case of perturbations from Jupiter, for cometary orbits with semi-major axes within the limits 10-100 AU and perihelion distances from 6 to 10 AU, show that the amplitudes of variations of inclinations are small and reach only 2-3° and that the periods of rotation of the line of apsides are measured by tens of thousands, hundreds of thousands and millions of years. The variation of the inclination angle can have no special cosmological significance. The rotation of the line of apsides causes a periodic variation of the smallest distances to which comets can approach a planetary orbit.

These distances change within the limits $q - R$ to $p - R$, where q, p are the perihelion distance and the parameter of the cometary orbit, R is the radius of the circular orbit of the planet. As a result of the variations of these distances, comets can evidently sometimes approach close to planets and be transformed into short-period comets.

1. Introduction

Several investigators have established [1-3] that the solar system includes a huge number of comets moving at great distances from the sun. Such comets are inaccessible to observation and can exist practically indefinitely since their great distance from the sun enables them to retain their constituent gas.

A comet can be observed only when the perihelion distance of its orbit is reduced so that it falls into the region of visibility.

The orbits of comets in the zone outside of Neptune are continuously changing due to perturbations by the planets. This is evidently the reason for the appearance of new short-period comets.

In the present paper we shall endeavor to explain how cometary orbits change over a long period of time due to planetary perturbations and how this leads to the appearance of observable short-period comets.

2. Planetary Perturbations

We obtain an approximate solution for the changes of cometary orbits under planetary perturbations by using the doubly averaged limit circular problem of three points.

The differential equations of this problem have the following integrals [4]:

$$a = \text{const}, \tag{1}$$

$$\sqrt{P} \cos i = \text{const}. \tag{2}$$

For the purpose of determining how the angle of inclination i changes along with the orbital parameter P, let us consider the following pair of differential equations [5]:

$$\frac{di}{dt} = \frac{1}{4\pi^3 k \sqrt{P}} \int_0^{2\pi}\int_0^{2\pi} rW \cos U \, dM \, dM', \tag{3}$$

$$e\frac{d\omega}{dt} = \frac{1}{4\pi^2 k \sqrt{P}} \int_0^{2\pi}\int_0^{2\pi} [-P \cos VS + (r+P)\sin VT - re \cot i \sin UW] \, dM \, dM'. \tag{4}$$

Here r, P, e, ω, i, V, M, U denote, respectively, the radius vector, parameter, eccentricity, angular distance of the perihelion from the node, inclination, true anomaly, mean anomaly and cometary latitude; M' denotes the mean anomaly of the planet; S, T, W are the components of the perturbing acceleration along the radius vector of the comet, in a direction perpendicular to the radius vector and lying in the plane of the osculating orbit, and along the normal to the orbital plane.

We represent S, T, and W as follows:

$$S = S_1 - S_2; \tag{5}$$

$$T = T_1 - T_2; \tag{6}$$

$$W = W_1 - W_2, \tag{7}$$

$$\left.\begin{aligned} S_1 &= \left(\frac{k^2m}{\Delta^3}\right)_S; \quad S_2 = \left(\frac{k^2m}{R^3}\right)_S; \\ T_1 &= \left(\frac{k^2m}{\Delta^3}\right)_T; \quad T_2 = \left(\frac{k^2m}{R^3}\right)_T; \\ W_1 &= \left(\frac{k^2m}{\Delta^3}\right)_W; \quad W_2 = \left(\frac{k^2m}{R^3}\right)_W. \end{aligned}\right\} \tag{8}$$

Here Δ is the distance between the comet and planet, m is the mass of the perturbing planet and R is the radius of the circular planetary orbit in the plane of the ecliptic.

From Figure 1 it follows that

$$S_1 = \frac{k^2m\,(r - R\cos L)}{\Delta^3}; \tag{9}$$

$$T_1 = -\frac{k^2mR}{\Delta^4}(\sin U \cos M' - \cos U \sin M' \cos i); \tag{10}$$

$$W_1 = \frac{k^2mR}{\Delta^3} \sin M' \sin i. \tag{11}$$

Here L is the angle between R and r.

From simple geometrical considerations we find that

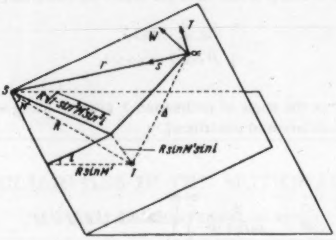

Fig. 1.

$$S_2 = \frac{k^2 m}{R^2} (\cos U \cos M' + \sin U \sin M' \cos i); \tag{12}$$

$$T_2 = -\frac{k^2 m}{R^2} (\sin U \cos M' - \cos U \sin M' \cos i); \tag{13}$$

$$W_2 = -\frac{k^2 m}{R^2} \sin M' \sin i. \tag{14}$$

Substituting in Equations (5), (6) and (7) instead of S_1, T_1, W_1 and S_2, T_2, W_2 their values from (9)-(14), we obtain

$$S = \frac{k^2 m\,(r - R\cos L)}{\Delta^3} - \frac{k^2 m}{R^2} (\cos U \cos M' + \sin U \sin M \cos i); \tag{15}$$

$$T = k^2 m\,R\left(\frac{1}{R^3} - \frac{1}{\Delta^3}\right)(\sin U \cos M' - \cos U \sin M' \cos i); \tag{16}$$

$$W = \frac{k^2 m R}{\Delta^3} \sin M' \sin i + \frac{k^2 m}{R^2} \sin M' \sin i. \tag{17}$$

Substituting these values of S, T, W into the differential equations (3) and (4), we have

$$\frac{di}{dt} = \frac{k^2 m}{4\pi^2 k\sqrt{P}} \int_0^{2\pi}\!\!\int_0^{2\pi} \left[\frac{R \sin M' \sin i}{\Delta^3} + \frac{\sin M' \sin i}{R^2}\right] r \cos U\, dM'\, dM; \tag{18}$$

$$\begin{aligned} e\frac{d\omega}{dt} = \frac{k^2 m}{4\pi^2 k\sqrt{P}} \int_0^{2\pi}\!\!\int_0^{2\pi} \Bigg\{ &- P\cos V\left[\frac{r - R\cos L}{\Delta^3} - \frac{1}{R^2}(\cos U \cos M' + \right. \\ &\left. + \sin U \sin M' \cos i)\right] + (r + P)\sin V\left[-\frac{R}{\Delta^3}(\sin U \cos M' - \right. \\ &\left. - \cos U \sin M' \cos i) + \frac{1}{R^2}(\sin U \cos M' - \cos U \sin M' \cos i)\right] - \\ &- re\cot\, i - \sin U\left[\frac{R}{\Delta^3}\sin M' \sin i + \frac{1}{R}\sin M' \sin i\right]\Bigg\}\, dM\, dM'. \end{aligned} \tag{19}$$

Considering that through integration the second term in (18), and the second, fourth and sixth terms in (19) vanish, and also transforming by means of the formula

$$dM = \frac{r^2}{a^{3/2}\sqrt{P}}\, dV \tag{20}$$

from dM to dV, we obtain

$$\frac{di}{dt} = -\frac{km\, R \sin i}{4\pi^2 Pa^{3/2}} \int_0^{2\pi}\!\!\int_0^{2\pi} \frac{r^3 \cos U \sin M'}{\Delta^3}\, dV\, dM'; \tag{21}$$

$$e\frac{d\omega}{dt} = \frac{km}{4\pi^3 Pa^{3/2}} \int_0^{2\pi}\!\!\int_0^{2\pi} r^2 \left\{ \frac{-P\cos V\,(r - R\cos L)}{\Delta^3} - \right.$$

$$\left. - \frac{(r+P)\sin V R\,(\sin U \cos M' - \cos U \sin M' \cos i)}{\Delta^3} - \frac{Rer\,\mathrm{ctg}\, i \sin U \sin i}{\Delta^3} \right\} dV\, dM'. \tag{22}$$

It is easily seen that $\Delta^2 = R^2 + r^2 - 2Rr\cos L$. We now find $1/\Delta^3$.

$$\frac{1}{\Delta^3} = \frac{1}{(R^2 + r^2 - 2Rr\cos L)^{3/2}} - \frac{1}{(R^2 + r^2)^{3/2}}\left[1 - \frac{2Rr\cos L}{R^2 + r^2}\right]^{-3/2} =$$

$$= \frac{1}{(R^2 + r^2)^{3/2}}\left[1 + \frac{3Rr\cos L}{K^2 + r^2} + \ldots\right] = \frac{1}{(R^2 + r^2)^{3/2}} + \frac{3Rr\cos L}{(R^2 + r^2)^{5/2}} + \ldots \tag{23}$$

The last terms in the expansion (23) are small enough to be neglected.

Substituting into (21) and (22) instead of $1/\Delta^3$ its value from (23), we obtain

$$\frac{di}{dt} = \frac{kmR\sin i}{4\pi^2 Pa^{3/2}} \int_0^{2\pi}\!\!\int_0^{2\pi} r^3 \cos U \sin M' \left[\frac{1}{(R^2 + r^2)^{3/2}} + \frac{3Rr\cos L}{(R^2 + r^2)^{5/2}}\right] dM\, dM; \tag{24}$$

$$e\frac{d\omega}{dt} = \frac{km}{4\pi^2 Pa^{3/2}} \int_0^{2\pi}\!\!\int_0^{2\pi} r^2 \left[\frac{1}{(R^2 + r^2]^{3/2}} + \frac{3Rr\cos L}{(R^2 + r^2)^{5/2}}\right] \times$$

$$\times [-P\cos V\,(r - R\cos L) - (r+P)\sin V R\,(\sin U \cos M' -$$

$$- \cos U \sin M' \cos i) - Rer\,\mathrm{ctg}\, i \sin U \sin i]\, dV \cdot dM'. \tag{25}$$

In order to simplify (24) and (25) we expand $\frac{1}{(R^2 + r^2)^{3/2}}$ and $\frac{1}{(R^2 + r^2)^{5/2}}$ in powers of $\frac{R^2}{r^2}$:

$$\frac{1}{(R^2 + r^2)^{3/2}} = \frac{1}{r^3}\left(1 + \frac{R^2}{r^2}\right)^{-3/2} = \frac{1}{r^3}\left[1 - \frac{3}{2}\frac{R^2}{r^2} + \frac{15}{8}\frac{R^4}{r^4} + \ldots\right] =$$

$$= \frac{1}{r^3} - \frac{3}{2}\frac{R^2}{r^5} + \frac{15}{8}\frac{R^4}{r^7} + \ldots; \tag{26}$$

$$\frac{1}{(R^2 + r^2)^{5/2}} = \frac{1}{r^5}\left(1 - \frac{R^2}{r^2}\right)^{-5/2} = \frac{1}{r^5}\left[1 - \frac{5}{2}\frac{R^2}{r^2} + \frac{35}{8}\frac{R^3}{r^4} + \ldots\right] =$$

$$= \frac{1}{r^5} - \frac{5}{2}\frac{R^2}{r^7} + \frac{35}{8}\frac{R^4}{r^9} + \ldots \tag{27}$$

Since we are considering only those comets for which we always have r > R, we can confine ourselves to only the first terms in these series.

It is easily seen that the following simple relations exist:

$$\cos L = \cos U \cos M' + \sin U \sin M' \cos i; \tag{28}$$

$$\cos U = \cos V \cos\omega - \sin V \sin\omega; \tag{29}$$

$$\sin U = \sin V \cos\omega + \cos V \sin\omega. \tag{30}$$

We substitute in (24) and (25) instead of

$$\frac{1}{(R^2 + r^2)^{3/2}} \text{ and } \frac{1}{(R^2 + r^2)^{5/2}}, \ \cos L, \ \cos U, \ \sin U$$

their values from (26)-(30). After integration and a simple transformation we obtain these equations in the form

$$\frac{di}{dt} = -\frac{45kmR^4}{64a^{3/2}p^4}(a - P)\sin 2i \sin 2\omega; \tag{31}$$

$$\frac{d\omega}{dt} = \frac{3}{8}\frac{kmR^2}{a^{3/2}p^3}(3.5\cos^2 i - \sin^2 i \cos^2\omega - 4.5]. \tag{32}$$

For more convenient integration of these equations we transform them through introduction of the quantities α, $\bar{P}$, τ, A, f and W by means of the following relations:

$$\alpha = \frac{R}{a}; \tag{33}$$

$$P = a\bar{P} = \frac{R}{\alpha}\bar{P}; \tag{34}$$

$$dt = \frac{R^{3/2}}{km\alpha^{7/2}}d\tau, \tag{35}$$

from which

$$t - t_0 = \frac{R^{3/2}}{km\alpha^{7/2}}\tau,$$

$$\sqrt{P}\cos i = A, \tag{36}$$

from which

$$\bar{P} = \frac{A^2}{\cos^2 i}; \tag{37}$$

$$\cos i = f; \tag{38}$$

$$2\omega = W. \tag{39}$$

Using these formulas, we transform (31) and (32) to

$$\frac{df}{d\tau} = \frac{45}{32}\frac{\alpha^3}{A^8}f^7(f^2 - A^2)(1 - f^2)\sin W; \tag{40}$$

$$\frac{dW}{d\tau} = \frac{3}{4}\frac{f^4}{A^4}\left[4(1+f^2) - \frac{1}{2}(1-f^2)\cos W\right]. \tag{41}$$

We shall seek the solution of this set of equations as series of powers of the small quantity α:

$$f = f_0 + \alpha^2 f_2 + \alpha^4 f_4 + \ldots; \tag{42}$$

$$W = W_0 + \alpha^2 W_2 + \alpha^4 W_4 + \ldots. \tag{43}$$

Differentiating (42) and (43) and substituting the values of these derivatives into the left-hand sides of (40) and (41), and also substituting into the right-hand sides of these equations instead of f and W their values from (42) and (43) and equating coefficients of equal powers of α, we obtain

$$\frac{df_0}{d\tau} = 0, \text{ i. e., } f_0 = \text{const}; \tag{44}$$

$$\frac{dW_0}{d\tau} = \frac{3}{4}\frac{f_0^4}{A^4}\left[4(1+f_0^2) - \frac{1}{2}(1-f_0^2)\cos W_0\right]. \tag{45}$$

Transforming (45), we obtain

$$\frac{dW_0}{d\tau} = \frac{3}{4}\frac{f_0^4}{A^4}4(1+f_0^2)\left[1 - \frac{1}{8}\frac{1-f_0^2}{1+f_0^2}\cos W_0\right]. \tag{46}$$

We introduce the notation

$$3\frac{f_0^4}{A^4}(1+f_0^2) = B \tag{47}$$

and

$$\frac{1}{8}\frac{1-f_0^2}{1+f_0^2} = n. \tag{48}$$

Substituting B and n into (46), we have

$$\frac{dW_0}{d\tau} = B(1 - n\cos W_0) \tag{49}$$

For the determination of f_2 we obtain

$$\frac{df_2}{d\tau} = \frac{45}{32}\frac{f_0^7}{A^8}(f_0^2 - A^2)(1 - f_0^2)\sin W_0. \tag{50}$$

To determine f_2 we perform the following transformations.

We can write

$$\frac{df_2}{d\tau} = \frac{df_2}{dW_0}\frac{dW_0}{d\tau}. \tag{51}$$

Using (34), (37), (47), (48), (49), (50), Equation (51) can be represented as

$$\frac{df_2}{dW_0} = \frac{15n}{4}\frac{f_0 a}{P^2}(a - P)\frac{\sin W_0}{1 - n\cos W}. \tag{52}$$

By integrating we obtain

$$f_2 = \frac{15af_0(a - P)}{4P^2}\ln(1 - n\cos W_0) + \text{const.} \tag{53}$$

Therefore, neglecting terms containing higher powers of α, *f* can be represented in the form

$$f = f_0 + \alpha^2\frac{15af_0(a - P)}{4P^2}\ln(1 - n\cos W_0). \tag{54}$$

This equation shows directly that *f* is a periodic function of W_0. If in the expansion (43) we neglect terms containing the small quantity α^2 as a factor, we shall have the equality

$$W_0 = 2\omega;$$

from which it follows that *f* is a periodic function of ω. Equation (54) shows that *f* has a minimum value at $W_0 = 0$, that is, at $\omega = 0$ and a maximum at $W_0 = \pi$, that is, at $\omega = \pi/2$.

From these relations we obtain the formula for calculating the amplitude (am) of *f* in the form

$$amf = \alpha^2\frac{15af_0(a - P)}{4P^2}\ln\frac{1 + n}{1 - n}. \tag{55}$$

In calculating amf from this formula we shall assume that for $t = 0$, $W_0 = \pi/2$ and thus $f = f_0$, that is f_0 is the initial value of *f*.

Equation (55) was used to calculate amf from orbits with different semi-major axes, different parameters and different inclinations in the case of perturbations by Jupiter. These calculations showed that the amplitudes of *f* are so small that the corresponding changes of inclination *i* amount to only 2° and in a few instances 3°. It follows from (2) that with a change of *i* there will also be a change of the orbital parameter P, and that a change of the parameter is accompanied by a change of the perihelion distance of the cometary orbit. This change is small and does not exceed 0.2 AU. Such changes are so insignificant that they cannot play an important part in the transformation of cometary orbits.

We shall now consider the result of the rotation of the line of apsides of a cometary orbit. We shall begin by deriving a formula for calculating the period of this rotation.

Since *f* changes very little with time, we shall assume in (41) that *f* is constant and equal to f_0.

Then, using (47) and (48), we represent (41) as

$$Bd\tau = \frac{dW}{1 - n\cos W}, \tag{56}$$

from which

$$B\tau = \int_0^W \frac{dW}{1 - n\cos W}. \tag{57}$$

For convenience in integrating we introduce the new variable σ which is associated with changes of W by the relation

$$\tan\frac{\sigma}{2} = \sqrt{\frac{1+n}{1-n}}\tan\frac{W}{2}. \tag{58}$$

It is easily seen that we then have the following relations:

$$1 - n\cos W = \frac{1-n^2}{1+n\cos\sigma}; \tag{59}$$

$$dW = \frac{\sqrt{1-n^2}}{1+n\cos\sigma}\,d\sigma. \tag{60}$$

Substituting in (57) instead of 1 − n cos W and dW their values from (59) and (60), we have

$$B\tau = \frac{1}{\sqrt{1-n^2}}\int_0^\sigma d\sigma. \tag{61}$$

from which

$$B\tau = \frac{\sigma}{\sqrt{1-n^2}}. \tag{62}$$

We shall assume that when σ is increased by 2π, τ changes by τ', and shall therefore write

$$B(\tau+\tau') = \frac{\sigma+2\pi}{\sqrt{1-n^2}}. \tag{63}$$

From (63) we now subtract (62) and obtain

$$\tau' = \frac{2\pi}{B\sqrt{1-n^2}}. \tag{64}$$

We now go over from τ to $\underline{t}$.

When τ varies by the amount τ', let $\underline{t}$ vary by the amount t'. Then, using (36), we obtain

$$t' = \frac{R^{3/2}}{kma^{1/2}}\,\tau'. \tag{65}$$

Using (64) and (47), we obtain a formula for calculating the time interval during which ω changes by π, in the form

$$t' = \frac{2\pi a^{7/2}A^4}{3f_0^4R^2mk(1+f_0^2)\sqrt{1-n^2}}. \tag{66}$$

Taking the solar mass as unity, the astronomical unit as unit of length and setting $\underline{k}$ equal to 2π, t' is expressed in years.

By using this result as well as some of Equations (33)-(39), we can represent (66) in the form

$$t' = \frac{a^{3/2}p^{2}}{3mR^{2}(1+f_0^{2})\sqrt{1-n^{2}}}\ . \tag{67}$$

This formula has been used to calculate the periods of variation of the angular separation of the perihelion from the node which results from perturbations by Jupiter. The calculations were performed for orbits with semi-major axes from 10 to 100 AU at intervals of 10 AU, inclinations from 5 to 30° at intervals of 5° and for perihelion distances from 6 to 10 AU at intervals of 1 AU.

The results are shown in Tables 1-6.

The tables show the periods of variation of the angular distances of the perihelion from the node for cometary orbits with different semi-major axes, inclinations and perihelion distances.

In order to obtain the period of variation of the angular distances of the perihelion from the node in years, each value in Tables 1-6 must be multiplied by 10^4. These tables have been used for the plotting of curves showing the dependence of the periods of change of ω on the semi-major axis for different angles of inclination and different perihelion distances.

The curves for the dependence of the period of variation of ω on a for different i and q are shown in Figures 2-7.

TABLE 1-2

For $i = 5°$						For $i = 10°$					
a \ q	6	7	8	9	10	a \ q	6	7	8	9	10
10	3	3	4	4	—	10	3	3	4	4	—
20	12	15	19	23	26	20	12	16	19	23	26
30	25	30	41	50	59	30	25	30	41	50	60
40	40	53	68	85	100	40	41	54	69	86	101
50	58	77	99	123	148	50	59	78	100	124	150
60	78	104	134	167	202	60	79	106	136	168	204
70	100	134	172	215	262	70	101	136	174	217	265
80	123	166	214	267	325	80	124	168	216	270	329
90	148	200	258	323	394	90	150	202	261	327	399
100	175	236	305	382	469	100	177	240	309	387	474

TABLE 3-4

For $i = 15°$						For $i = 20°$					
a \ q	6	7	8	9	10	a \ q	6	7	8	9	10
10	3	3	4	4	—	10	3	4	4	4	—
20	13	16	20	23	27	20	13	16	20	24	28
30	26	31	42	51	61	30	26	32	43	53	63
40	42	55	70	87	103	40	43	57	72	90	106
50	60	80	102	126	153	50	62	82	105	130	157
60	81	108	138	172	208	60	83	111	142	177	214
70	103	138	178	227	270	70	106	142	183	228	277
80	127	171	221	275	336	80	131	176	227	283	345
90	154	206	266	333	406	90	158	212	274	342	418
100	181	244	314	394	483	100	186	256	324	405	495

TABLE 5-6

For $i = 25°$						For $i = 30°$					
q \ a	6	7	8	9	10	q \ a	6	7	8	9	10
10	3	4	4	4	—	10	3	4	4	4	—
20	13	17	21	25	29	20	14	18	22	26	30
30	27	33	45	54	65	30	28	35	46	57	67
40	44	58	74	92	110	40	46	61	77	96	114
50	64	85	108	134	162	50	66	89	113	140	169
60	86	114	147	182	221	60	89	119	153	190	230
70	109	147	189	235	286	70	113	153	196	245	298
80	135	182	234	292	356	80	140	189	244	304	371
90	163	219	283	353	431	90	169	226	294	363	449
100	192	258	334	418	511	100	199	269	348	436	532

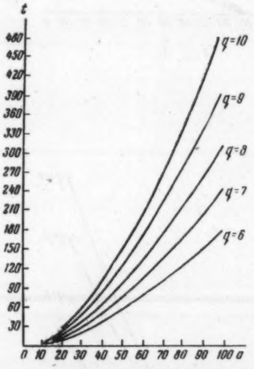

Fig. 2.

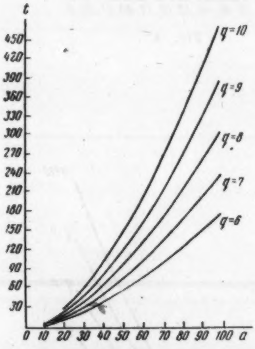

Fig. 3.

The data in the tables and the curves show that the angular distance of the perihelion from the node of cometary orbits varies from 0 to 2π in tens of thousands, hundreds of thousands and even millions of years.

It is easily seen that the rotation of the lines of apsides of cometary orbits results in a change of the minimum distance of approach of the comets to planetary orbits. These distances will be smallest in the two cases $\omega = 0°$ and $\omega = 180°$. In these two cases the smallest distance of approach by a comet to a planetary orbit is $q - R$, where $\underline{q}$ is the perihelion distance of the cometary orbit and R is the radius of the circular planetary orbit.

This distance will be greatest when $\omega = 90°$ or $\omega = 270°$. It will be greatest for comets moving in orbits inclined at 90°. In this case the distance is $P - R$, where P is the parameter of the cometary orbit and R has its previous value.

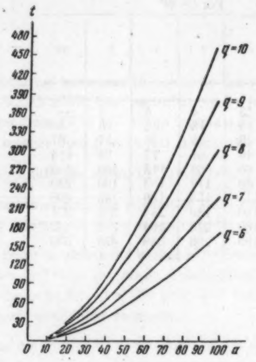

Fig. 4.

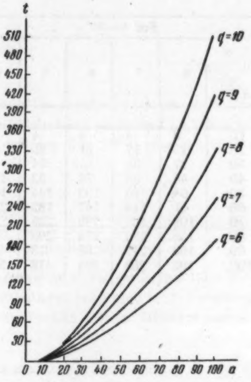

Fig. 5.

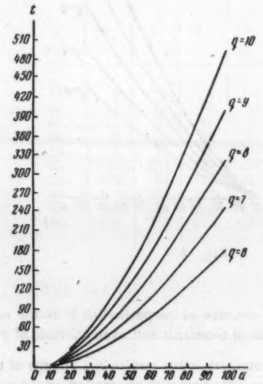

Fig. 6.

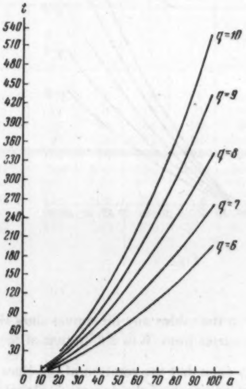

Fig. 7.

The changes of these distances can evidently result in close approach of comets and planets; this can result in a radical transformation of cometary orbits and the appearance of short-period comets.

In conclusion I wish to thank A. A. Orlov for suggestions.

Kursk State Pedagogical Institute

Received June 9, 1956

LITERATURE CITED

[1] V. A. Krat, Voprosy Kosmogonii (Problems of Cosmogony), (Izd. AN SSSR 1, 34, 1952).*

[2] V. G. Fesenkov, Astron. J. (USSR) 28, 98 (1951).

[3] V. Oort, Bull. Astronom. Inst. Netherlands 13, 91 (1950).

[4] N. D. Moiseev, Trudy GAISh 25, Bk. 1, 75 (1945).

[5] M. F. Subbotin, Course in Celestial Mechanics II (ONTI, 1937).*

---

*In Russian.

THE DISTRIBUTION OF COMETS OF JUPITER'S FAMILY

K. A. Shteins

A two dimensional function of the distribution of semi-major axes and parameters of the orbits of comets, captured by Jupiter, is determined by the generally accepted method. The distribution of long-period comets, disintegration and the probability of discovery of short-period comets are taken into account. It is assumed that the probability of discovery of short and long-period comets is the same and determined from observations empirically. According to observational data it is assumed that the distribution of the perihelia of long-period comets is proportional to the distance of the perihelia from the sun. The influence of the disintegration of comets is investigated in accordance with suggestions of S. Orlov and B. Levin and the gravitational hypothesis. The results are in good agreement with observations. The almost complete absence of short-period comets with retrograde motions is a result of the fact that these comets, in most cases, have small perihelion distances and disintegrate rapidly or, to a lesser extent, that they are transformed from a comparatively small number of long-period comets, which have small perihelion distances.

The Plane Problem

§1. The existence of a group of short-period comets whose aphelia are close to Jupiter's orbit, shows that Jupiter has played an important part in their origin and perhaps in the origin of other short-period comets. Investigations by Tisserand, Schulhoff and others [1] have shown that some of the characteristics of Jupiter's comet family indicate that this group of comets originated through capture by Jupiter of some long-period comets. H. A. Newton [1] investigated the theoretical distribution of short-period comets and showed that direct motion must predominate over retrograde motion among short-period comets. Newton's results cannot be compared exactly with observations because he considered one-dimensional distributions. The majority of scientists have regarded Newton's investigations as confirming the hypothesis that short-period comets originated through capture; but others, like S. K. Vsekhsviatskii [2], have regarded these same investigations as disproving this possibility. Newton's calculations show that comets whose planes have small inclinations should arise with direct and retrograde motion in about the same numbers. In actuality, the majority of short-period comets have small inclinations and are not observed to have retrograde motion. B. M. Shchigolev [3] has made a special study of the orbit elements of periodic comets captured by Jupiter. He considered the plane problem for the hypotheses of Newton and of Fesenkov. According to Newton's hypothesis the velocities of long-period (parabolic) comets are distributed uniformly at each point of interplanetary space, and according to Fesenkov's hypothesis the velocities of parabolic comets outside the sun's sphere of influence can have completely arbitrary orientations. Fesenkov's and Newton's hypotheses were used for ease of calculation and are, first of all, the source of the radical discrepancy between Shchigolev's results, which gave approximately identical probability for direct and retrograde motions, and our results, according to which direct motions are almost the only possible motions. Secondly, Shchigolev used Newton's approximate formulas, thus introducing another small difference in the results. Thirdly, Shchigolev did not treat the matter of comet disintegration; in 1924, when his article was written, this question had not arisen. It must be emphasized that Shchigolev's work shows that an investigation of direct versus retrograde motions requires great caution. His investigation shows that Fesenkov's and Newton's hypotheses lead to different results.

In the present article we again investigate the distribution of the comets in Jupiter's family, but we take into account the distribution of parabolic cometary orbits, the lifetimes of short-period comets and the probability of discovering short-period comets. We assume that we have a plane problem because the inclination of the orbits of Jupiter's family is small. This is not entirely legitimate. Although all the members of the family of short-period comets possess small inclinations, in the original family of long-period comets arbitrary inclinations are encountered. In the second part of the present article we shall consider the three-dimensional problem, and shall determine the order of the error introduced by our assumption. Since the three-dimensional problem leads to fourfold integrals, the modest calculational resources of the author will cause it to be treated less precisely than the plane problem. We have therefore considered it necessary to divide the work and to consider the plane problem separately. We propose subsequently to examine the problem of double capture of comets and similar questions.

In its general features the scheme of this investigation is classical. We assume that a stationary flock of parabolic comets is revolving around the sun [4]. Stellar perturbations [5] insure the stationary condition. Perturbation by Jupiter changes some of the flock of parabolic comets into short-period comets. We assume that Jupiter has a circular orbit of radius R_0 around the sun. Jupiter does not influence the motion of a comet at a large distance, but inside its circle of influence only Jupiter attracts the comet. Dynamic effects of the rotation of the circle of influence are not taken into account. As a result of the assumed symmetry the passage of a parabolic comet through Jupiter's circle of influence is considered for only one definite position of Jupiter.

We shall describe the path of a comet which undergoes capture. The orbit of the parabolic comet is determined by the perihelion distance $\underline{q}$ and perihelial longitude ω. The longitude of the perihelion is computed from the straight line CJ which connects the sun (C) and Jupiter (J). The position of the comet in its orbit is not fixed, since it is assumed that a continuous flow of comets takes place along the orbit in both directions. The comet intersects the circle of influence at the point $K(\alpha, \rho)$, where α is the angle in Jupiter's polar coordinates with the pole J and polar axis CJ. The radius of Jupiter's circle of influence is denoted by ρ. The angle between the comet's velocity $\mathbf{v}_k$ and CJ is denoted by β. The comet enters the circle of influence with the relative velocity $\mathbf{v}_r$, which is determined by the coordinates α, β_1. To determine the motion of the comet inside the circle of influence we introduce the coordinates (α, β_2), where $\beta_2 = \beta_1 - \alpha$ is the angle between the normal of the circle of influence at the point α and the relative velocity. Inside the circle of influence the comet describes a hyperbola and leaves the circle at the point $\alpha + 2\varphi$. The relative velocity of the comet $\mathbf{v}'_r (v_r = v'_r)$ at this point forms the angle $\pi - \beta_2$ with the normal. The absolute velocity $\mathbf{v}_a = \mathbf{v}'_r + \mathbf{v}_J$, where $\mathbf{v}_J$ is the velocity of Jupiter, forms the angle β_3 with CJ. The elliptic orbit of the captured comet is determined by the semi-major axis $\underline{a}$ and the parameter $\underline{p}$.

The problem of the present article is the determination of the theoretical distribution f(a, p) of captured comets. The problem will be solved in the following form:

$$\int | f(a, p) | \, da \, dp =$$
$$= \int \left| f_1(q, \omega) \frac{D(q, \omega)}{D(\alpha, \beta)} \frac{D(\alpha, \beta)}{D(\alpha + \beta_2, \beta_2)} v_r \cos \beta_2 \frac{D(\alpha + \beta_2, \beta_2)}{D(a, p)} f_2(a, p) f_3(a, p) \right| da \, dp. \qquad (1)$$

Here $f_1(q, \omega)$ is the distribution function of parabolic comets, $f_2(a, p)$ is the probability of the lifetime of short-period comets and $f_3(a, p)$ is the distribution function of the discovery of comets. The Jacobian

$\left| \frac{D(\alpha + \beta_2, \beta_2)}{D(a, p)} \right|$ is defined as the area between the curves $a = c_i$, $p = c'_j$ in the rectangular coordinates $\alpha + \beta_2$, β_2. This procedure is simplest and the limits of integration are obtained directly.

§2. From ΔCKJ we have

$$(\mathrm{CK})^2 = R_0^2 + \rho^2 + 2R_0\rho\cos\alpha, \tag{2}$$

$$\sin\measuredangle \mathrm{KCJ} = \frac{\rho}{(\mathrm{CK})}\sin\sigma. \tag{3}$$

We shall regard $\rho/R_0 \approx 0.06$ as a first-order small quantity and shall hereinafter restrict ourselves to only first-order quantities. Therefore

$$\mathrm{CK} = R_0\left[1 + \frac{\rho}{R_0}\cos\alpha\right], \tag{4}$$

$$\measuredangle \mathrm{KCJ} = \frac{\rho}{R_0}\sin\alpha. \tag{5}$$

The parabola intersects Jupiter's circle of influence at two points. We shall consider each parabola twice, because comets move along parabolas in both the direct and retrograde directions. Then one definite pair of values of α, β corresponds to each parabola. Directions on the parabola are chosen so that in relative motion the comets travel inside of the circle of influence.

On the basis of the familar parabolic property that the angle between the parabolic axis and the radius vector is twice the angle between the tangent and the radius vector we easily obtain the following equality:

$$\omega = 2\beta - \pi - \measuredangle \mathrm{KCJ}. \tag{6}$$

Equations (5) and (6) give

$$\omega = 2\beta - \pi - \frac{\rho}{R_0}\sin\alpha. \tag{7}$$

Since

$$\mathrm{CK} = \frac{2q}{1 + \cos(\omega - \measuredangle \mathrm{KCJ})}, \tag{8}$$

we have

$$q = \frac{R_0}{2}\left\{1 - \cos 2\beta + (1 - \cos 2\beta)\frac{\rho}{R_0}\cos\alpha - 2\sin 2\beta\,\frac{\rho}{R_0}\sin\alpha\right\}. \tag{9}$$

From (7) and (9) we easily obtain

$$\left|\frac{D(q,\omega)}{D(\alpha,\beta)}\right| = 2\rho\,|\sin\beta\cos(\alpha - \beta)|. \tag{10}$$

The comet meets the circle of influence with the absolute velocity

$$v_H = \sqrt{2}\,v_J\left[1 - \frac{\rho}{2R_0}\cos\alpha\right]. \tag{11}$$

Inside of the circle of influence the comet travels with the relative velocity $\mathbf{v}_r = \mathbf{v}_k - \mathbf{v}_J$; therefore

$$v_r \cos\beta_1 = v_r \cos(\alpha + \beta_2) = v_K \cos\beta; \tag{12}$$

$$v_r \sin\beta_1 = v_r \sin(\alpha + \beta_2) = v_K \sin\beta - v_J. \tag{13}$$

It then follows that

$$v_r = v_J\left[-\sin(\alpha+\beta_2) + \sqrt{2 - 2\frac{\rho}{R_0}\cos\alpha - \cos^2(\alpha+\beta_2)}\right]. \tag{14}$$

Because of the smallness of $\rho/R_0 \approx 0.06$, $|\pi - \beta_2| < 9°$ for $a \leq 5.2$ we replace (14) by

$$v_r = v_J\left[-\sin(\alpha+\beta_2) + \sqrt{2 + 2\frac{\rho}{R_0}\cos(\alpha+\beta_2) - \cos^2(\alpha+\beta_2)}\right]. \tag{15}$$

In determining the Jacobian

$$\left|\frac{D(\alpha,\beta)}{D(\alpha+\beta_2,\beta_2)}\right| = \left|\frac{\partial\beta}{\partial(\alpha+\beta_2)} + \frac{\partial\beta}{\partial\beta_2}\right| \tag{16}$$

we shall confine ourselves to zero-order quantities. Then in (16) the term $\partial\beta/\partial\beta_2$ can be dropped, because $\beta_2 = (\alpha + \beta_2) - \alpha$ appears in the equation only with the factor ρ/R_0. Consequently

$$\left|\frac{D(\alpha,\beta)}{D(\alpha+\beta_2,\beta_2)}\right| = \frac{v_r}{v_J\sqrt{2 + 2\frac{\rho}{R_0}\cos(\alpha+\beta_2) - \cos^2(\alpha+\beta_2)}}. \tag{17}$$

On the basis of (12), (13) and (9) we obtain the following approximate formula:

$$q = \frac{R_0}{2}\left[1 + \frac{v_r}{v_J}\sin(\alpha+\beta_2)\right]^2. \tag{18}$$

This a very good representation of the exact solution because if in the derivation we also included terms with the factor ρ/R_0 we would have obtained the formula for q with the factor

$$\left[1 + \frac{v_r}{v_J}\sin(\alpha+\beta_2)\right].$$

Substituting into (10) the values taken from (12) and (13), we obtain

$$\left|\frac{D(q,\omega)}{D(\alpha,\beta)}\right| = \rho\left|\left[1 + \frac{v_r}{v_J}\sin(\alpha+\beta_2)\right]\left[\sin\alpha + \frac{v_r}{v_J}\cos\beta_2\right]\right| \approx$$

$$\approx \rho \left| \left[1 + \frac{v_r}{v_J} \sin(\alpha + \beta_2) \right] \left[\frac{v_r}{v_J} + \sin(\alpha + \beta_2) \right] \right| =$$

$$= \rho \sqrt{2 + 2 \frac{\rho}{R_0} \cos(\alpha + \beta_2) - \cos^2(\alpha + \beta_2)} \sqrt{\frac{2q}{R_0}}. \qquad (19)$$

When we take (17) into account we finally obtain

$$\left| \frac{D(q, \omega)}{D(\alpha + \beta_2, \beta_2)} \right| v_r = \frac{\sqrt{2}\ \rho v_r^2}{\sqrt{R_0}\ v_J} \sqrt{q}. \qquad (20)$$

§3. The angular coordinate of the comet which is moving around Jupiter is given by the formula for a conic section:

$$\rho = \frac{P}{1 + E \cos \varphi}, \qquad (21)$$

where P is the parameter and E is the linear eccentricity of the hyperbola. P is determined by the law of areas according to

$$P = \frac{\rho^2 v_r^2 \sin^2 \beta_2}{k^2 m'}, \qquad (22)$$

where m' is the mass of Jupiter and k^2 is the constant of attraction. For motion around Jupiter we also have conservation of energy given by

$$v_r^2 = k^2 m' \left[\frac{2}{\rho} - \frac{1}{A} \right], \qquad (23)$$

where A is the semi-major axis of the orbit and

$$E^2 = 1 - \frac{P}{A}. \qquad (24)$$

From (21), (22) and (23) we easily obtain*

$$\sin^2 \beta_2 = \frac{[2k^2 m' \operatorname{tg}^2 \varphi + \rho v_r^2] \pm \sqrt{[2k^2 m' \operatorname{tg}^2 \varphi + \rho v_r^2]^2 - 4k^4 m'^2 \operatorname{tg}^2 \varphi (1 + \operatorname{tg}^2 \varphi)}}{2 \rho v_r^2 (1 + \operatorname{tg}^2 \varphi)} \qquad (25)$$

or for values of 2φ which are very close to 360° the more convenient formula

$$\sin \beta_2 = \pm \frac{k^2 m' \operatorname{tg} \varphi}{\sqrt{\rho}\, v_r \sqrt{2 \operatorname{tg}^2 \varphi\, k^2 m' + \rho v_r^2}}. \qquad (26)$$

---

* 'tg' ≡ 'tan' — Publisher.

The comet leaves the circle of influence at the point $\alpha + 2\varphi$ and angle $\pi - \beta_2$ with the normal. The transition to fixed space obeys the following formulas:

$$v_a \cos\beta_3 = v_r \cos(\pi - \beta_2 + \alpha + 2\varphi); \tag{25'}$$

$$v_a \sin\beta_3 = v_r \sin(\pi - \beta_2 + \alpha + 2\varphi) + v_J; \tag{26'}$$

$$v_a^2 = v_r^2 + 2v_r v_J \sin(\pi - \beta_2 + \alpha + 2\varphi) + v_J^2. \tag{27'}$$

Taking as the radius vector from the sun to the comet for the exit point

$$r' = R_0\left[1 + \frac{\rho}{R_0}\cos(\alpha + 2\varphi)\right], \tag{28}$$

we have, according to the energy integral for captured comets and (27),

$$\frac{1}{a} - \frac{1}{R_0} = -\frac{2\rho}{R_0^2}\cos(\alpha + 2\varphi) - \frac{v_r^2 + 2v_J v_r \sin(\pi - \beta_2 + \alpha + 2\varphi)}{k^2}. \tag{29}$$

Considering the smallness of the factor $2\rho/R_0^2$, we can in the first term of the right-hand side of (29) replace $\alpha + 2\varphi$ by $\pi - \beta_2 + \alpha + 2\varphi$. We then obtain from (29):

$$v_r = -v_J \sin(\pi - \beta_2 + \alpha + 2\varphi) \pm$$

$$\pm\sqrt{v_J^2 \sin^2(\pi - \beta_2 + \alpha + 2\varphi) - \frac{2\rho k^2}{R_0^2}\cos(\pi - \beta_2 + \alpha + 2\varphi) - k^2\left(\frac{1}{a} - \frac{1}{R_0}\right)}. \tag{30}$$

Similarly, by the law of areas, for captured comets we have

$$v_r = \frac{\frac{k\sqrt{p}}{R_0} - v_J - \frac{\rho v_J}{R_0}\cos(\pi - \beta_2 + \alpha + 2\varphi)}{\sin(\pi - \beta_2 + \alpha + 2\varphi)}. \tag{31}$$

For direct motions $\sqrt{p} > 0$, and for retrograde motions $\sqrt{p} < 0$.

The curves $a(\alpha + \beta_2, \beta_2) = \text{const}$ and $\sqrt{p}(\alpha + \beta_2, \beta_2) = \text{const}$ were determined as follows. For $a = 2.6, 2.8, \ldots, 5.2$ and $\sqrt{p} = -2.4, -2.2, \ldots, +1.3, +1.4, \ldots +2.0, +2.2$ with definite values of $\pi - \beta_2 + \alpha + 2\varphi$, v_r was determined from (30) and (31). From a previously compiled table of v_r with the argument $\pi + \beta_2 + \alpha$ (15) the corresponding values of $\pi + \beta_2 + \alpha$ were determined. In first approximation 2φ was determined from the formula

$$2\varphi = (\pi - \beta_2 + \alpha + 2\varphi) - (\pi + \beta_2 + \alpha) + 2\beta_2 \approx (\pi - \beta_2 + \alpha + 2\varphi) - (\pi + \beta_2 + \alpha).$$

With approximate values of $(2\varphi, \pi + \alpha + \beta_2)$ Equation (25') was used to obtain approximate values of $\pi - \beta_2$. In the second approximation the differences of $2\beta_2$ from 360° were taken into account. The results of the calculations are shown in Figure 1. The horizontal axis represents $\pi + \alpha + \beta_2 = \alpha + \beta''$ and the vertical axis represents $\beta'' = \pi + \beta_2$ with opposite sign. The continuous curves represent $a = \text{const}$; the dashed curves represent $\sqrt{p} = \text{const}$. The author possessed a more detailed graph. The values of the Jacobian

$$\left|\frac{D(\alpha + \beta_2, \beta_2)}{D(a, p)}\right| da\, dp$$

are proportional to the areas in the figure. Thus the areas marked with crosses represent the number of comets with direct motion for which $3.2 < a < 3.6$; $0.6 < \sqrt{p} < 1.0$, and the dotted area is proportional to the number of captured comets with retrograde motions for which $2.6 < a < 2.8$; $0.4 < \sqrt{p} < 0.8$. This result was obtained on the basis of the assumption that $\alpha + \beta''$ and β'' are distributed uniformly and that there is no selection of observations nor disintegration of comets. By taking into account the other factors in (1) we can make the solution more precise.

§4. The lifetime of a comet depends on several physical factors. First, the evaporation of the gases increases cometary instability, and the loss of gas can cause a comet to lose its luminosity.

In a number of articles Vsekhviatskii stated that the brightness of a comet decreases monotonically from one appearance to the next, but this has been disputed. Some scientists believe that photometric data on comet brightness as well as the theory of cometary luminosity are too incomplete to allow such conclusions. There have been calculations of comet lifetimes assuming desorption but no adsorption of gases. S. V. Orlov [6] determined the number of N_1 of molecules lost by the nucleus in unit time. The average for comets is

$$N_1 = \frac{s}{r^4}, \tag{32}$$

where s is a constant and r is the distance from the sun to the comet. According to Orlov, in each passage a comet loses an identical quantity of its gases. Upon this hypothesis it is easy to calculate the life T of a comet. We obtain the following relation between T and the elements of the cometary orbit:

$$T = s_1 a^{3/2}\left[\int_0^{T_0} \frac{dt}{r^4}\right]^{-1} = s_1 k p^{5/2} a^{3/2}\left[\int_0^{\pi} (1 + e\cos v)^2\, dv\right]^{-1} = s_2 p^{5/2} a^{3/2}\,[2 + e^2]^{-1}. \tag{33}$$

Here s_1 and s_2 are constants.

B. Iu. Levin [7] has derived a formula for the loss of gases from the nucleus of a comet, based on the assumption that the temperature of the nucleus is inversely proportional to $\sqrt{r}$. He proposed that we consider the store of gases N of a comet to be diminishing in accordance with the formula

$$N = N_0 (1 - k_1)^{p_1}, \tag{34}$$

where p_1 is the number of revolutions of the comet, and k_1 is the fraction of its molecules lost in one revolution. Assuming k_1 to be small (which leads to Orlov's hypothesis), we have, to first-order small quantities,

$$T = s_3 (1 - e)^{-3/2} e^{L\sqrt{q_1}/R\cdot 350^\circ}, \tag{35}$$

where L is the heat of sorption, R is the absolute gas constant and q_1 is the perihelion distance of the comet. On the average $L(R\cdot 350^\circ)^{-1} = 10$.

Secondly, the life of a comet depends on the attractive force. Comets have been observed directly to disintegrate into several parts which subsequently disintegrated into a swarm of meteorites. It is reasonable to assume that the disintegration resulted from an attractive force proportional to dt/r^3. The overall disintegration P during one period of revolution T_0 as unit of time is given by

$$P = a^{-3/2}\int_0^{T_0} \frac{dt}{r^3} = s_4 a^{-3/2} p^{-3/2}, \tag{36}$$

where s_4 is a constant. If it is assumed that this "disintegration" is a continuous depletion of the comet, we then have

$$T = s_5 a^{3/2} p^{3/2}, \tag{37}$$

where s_5 is a constant.

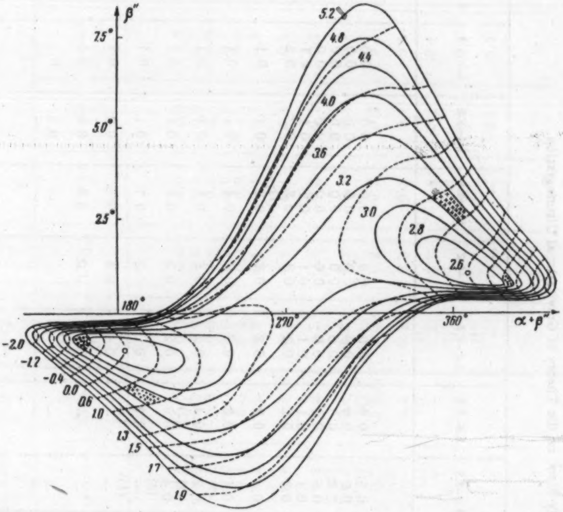

Fig. 1. Isoenergetic and isosectorial lines. The continuous lines are curves of identical values of the semi-major axes; the dashed lines represent identical parameters. Negative values represent comets with retrograde motions.

TABLE 1

q	f_1	q	f_1	q	f_1
0.0—0.3	1	1.0	0.43	1.8	0.056
0.4	0.83	1.2	0.24	2.0	0.048
0.6	0.73	1.4	0.12	$2.0-q$	$\frac{5.0-q}{\text{const}}$
0.8	0.58	1.6	0.074	—	

Since comets cease to emit light or disintegrate into swarms of meteorites we do not directly observe the distribution of Jupiter's comet family which results only from capture by Jupiter. In order to determine the observable distribution we must obtain the probability of the existence of a short-period comet. It is reasonable to assume that this probability is proportional to the lifetime of the comet. Then in accordance with the argument of this section we have three hypotheses for the function $f_2(a, p)$ corresponding to Formulas (33), (35) and (37).

§5. We have a well-developed theory on the probability of comet discovery $f_3(a, p)$. But, as N. D. Moiseev [8] has pointed out, this theory does not agree with observations. Therefore we shall determine $f_3(a, p)$ and $f_1(q, \omega)$ at the same time by an empirical method. For this purpose it is sufficient to assume that the

TABLE 2

Distribution of 26 Short-Period Comets of Jupiter's Family Based on the Theory of Gravitational Disintegration

a $\sqrt{p}$	2.6—2.8	2.8—3.0	3.0—3.2	3.4—3.4	3.4—3.6	3.6—3.8	3.8—4.0	4.0—4.2	4.2—4.4	4.4—4.6	4.6—4.8	4.8—5.0	5.0—5.2
0.0—0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2—0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.4—0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.6—0.8	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.8—1.0	0.1	0.8	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1.0—1.2	—	1.1	1.2	0.5	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
1.2—1.3	—	(1) —	(2) 2.0	0.7	0.3	0.2	0.2	0.2	(1) 0.2	0.1	0.1	0.1	0.1
1.3—1.4	—	(1) —	(1) 1.7	1.4	0.4	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1
1.4—1.5	—	—	(2) —	0.8	(1) 0.8	0.4	0.2	0.1	0.1	0.1	0.1	0.1	0.1
1.5—1.6	—	—	(2) —	(1) —	(1) 0.4	(2) 0.5	0.2	0.2	0.1	0.1	0.1	0.1	0.1
1.6—1.7	—	—	(1) —	—	(3) —	0.2	0.5	0.2	0.1	0.1	0.1	0.1	0.1
1.7—1.8	—	—	—	—	(1) —	(2) —	(1) 0.1	0.4	0.2	0.2	0.1	0.1	0.1
1.8—1.9	—	—	—	—	—	—	(1) —	0.0	0.3	0.4	0.1	0.1	0.1
1.9—2.0	—	—	—	—	—	—	(1) —	—	—	0.0	0.2	(1) 0.3	0.2
2.0—2.2	—	—	—	—	—	—	—	—	—	—	—	0.0	0.2

Total number of short-period comets with retrograde motion 0.17.

TABLE 3

Distribution of 26 Short-Period Comets of Jupiter's Family Based on Orlov's Theory

a $\sqrt{p}$	2.7–2.8	2.8–3.0	3.0–3.2	3.2–3.4	3.4–3.6	3.6–3.8	3.8–4.0	4.0–4.2	4.2–4.4	4.4–4.6	4.6–4.8	4.8–5.0	5.0–5.2
0.0–0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2–0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.4–0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.6–0.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.8–1.0	0.0	0.3	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0–1.2	—	0.7	0.7	0.3	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
1.2–1.3	—	(1) —	(2) 1.6	0.6	0.2	0.1	0.1	0.1	(1) 0.1	0.1	0.1	0.1	0.1
1.3–1.4	—	(1) —	(1) 1.7	1.4	0.4	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1
1.4–1.5	—	—	(2) —	1.0	(1) 0.9	0.4	0.2	0.1	0.1	0.1	0.1	0.1	0.1
1.5–1.6	—	—	(2) —	(1) —	(1) 0.5	(2) 0.7	0.3	0.2	0.2	0.2	0.1	0.1	0.1
1.6–1.7	—	—	(1) —	—	(3) —	0.4	0.7	0.3	0.2	0.1	0.1	0.1	0.1
1.7–1.8	—	—	—	—	(1) —	(2) —	(1) 0.2	0.7	0.4	0.2	0.1	0.2	0.2
1.8–1.9	—	—	—	—	—	—	(1) —	0.0	0.5	0.8	0.3	0.3	0.3
1.9–2.0	—	—	—	—	—	—	(1) —	—	—	0.1	0.5	(1) 0.7	0.4
2.0–2.2	—	—	—	—	—	—	—	—	—	—	—	0.1	0.6

Total number of short-period comets with retrograde motion 0.077.

discovery functions of short- and long-period comets are of the same form and that for very small q of long-period comets $f_3(a, p) \equiv 1$. This results from most theories on the probability of discovering comets. In this case for small q we see from Figure 2 that $f_1(q, \omega) = s_8 q$, where s_8 is a constant which is of no interest to us here. We take $f_3(a, p)$ directly from Figure 2 and obtain Table 1.

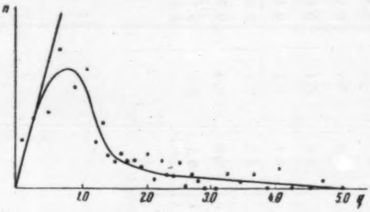

Fig. 2. Distribution of long-period comets discovered since 1700 as a function of perihelion distance q.

§6. From measurements of the elementary areas in Figure 1, from Equations (1), (14), (18), (20), (33), (35), (37) and from Table 1, assuming $f_1 = s_8 q$ we have obtained the results in Tables 2 and 3. The numbers within parentheses denote the numbers of detected short-period comets according to the catalogue of [9], for which $i < 10°$. Six comets were not taken into account because they cannot be included in Jupiter's family.

Tables 2 and 3 show directly that our theory, both in the form based on Orlov's hypothesis and in the variation based on gravitational disintegration, is in good agreement with observations. The fact that some of the discovered comets are somewhat outside of the limits predicted by theory, was explained by Tisserand [1]. The hypothesis based on Levin's proposition represents observations poorly.

As a form of analysis of our calculations we shall explain why we have obtained very small probabilities for comets with retrograde motion. This does not at all result from the fact that we did not use Newton's formulas but derived our own formulas (Section 3). The difference in the curves for $\beta^* > 0$ and $\beta^* < 0$ gives us an idea of the inaccuracy of Newton's formula, according to which the curves should be identical for $\beta^* > 0$ and $\beta^* < 0$. The existing discrepancy evidently cannot result in an essential difference in calculations of the probabilities.

The almost complete theoretical absence of short-period comets with retrograde motions results from the fact that these comets in most cases have small perihelion distances and disintegrate rapidly or, to a lesser extent, because they originate from among the relatively small number of long-period comets which have small perihelion distances.

Astronomical Observatory,
Latvian State University

Received June 19, 1956

LITERATURE CITED

[1] K. D. Pokrovskii, The Origin of Periodic Comets [in Russian] (1901)

[2] S. K. Vsekhsviatskii, Astron. J. (USSR) 32, 432 (1955).

[3] B. M. Shchigolev, Astron. J. (USSR) 1, 93 (1924).

[4] K. A. Shteins, Astron. J. (USSR) 30, 184 (1953).

[5] K A. Shteins, Astron. J. (USSR) 32, 282 (1955).

[6] S. V. Orlov, Astron. J. (USSR) 16, 3 (1939).

[7] B. Iu. Levin, Astron. J. (USSR) 20, 4, 37 (1943).

[8] N. D. Moiseev, Trudy GAISh (Shternberg State Astronomical Observatory) 6, 5 (1935).

[9] F. Baldet and G. de Obaldia, Catalogue Général des Orbites des Comètes de l' An 466 à 1952 (Paris, 1952).

ON THE COLLIMATION OF A TRANSIT INSTRUMENT

G. P. Pil'nik

It is shown that the collimation of a broken transit instrument varies from the beginning to the end of observations and also depends on the zenith distance During one hour of observations collimation varies on the average by $0^s.05$. Collimation differs by $0^s.2$ at the zenith and at zenith distance 55°. Therefore, the suggestion that collimation can be excluded by a reversal of the instrument during observations of each star is unacceptable.

For observations with a portable transit instrument it is assumed that by reversal of the instrument, collimation is totally excluded. Such an assumption is correct only if during observation the collimation remains constant. However, our investigations have established that the collimation changed throughout the observation and also with the zenith distance. Data on the variation of collimation with time were published for three transit instruments in the proceedings of the Shternberg State Astronomical Institute Vol. 27 (1956). Further studies of the instrument APM-1 No. 530001 (d = 100 mm, f = 100 cm) showed a considerable dependence of the collimation on the zenith distance.

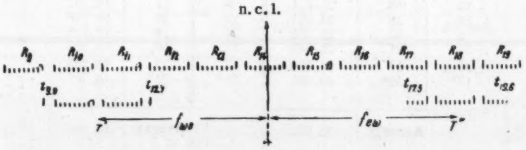

Fig. 1. Turns of micrometer screw of APM-1 No. 530001.

For the calculation of the collimation, we used running observations of the error of the clock. As the star passages were registered with identical turns of the micrometer screw, it is possible to calculate from the observations the mean distance of selected contacts from the noncollimated line (n. c. l.). In Figure 1 we see the contacts of part of the micrometer screw of APM-1 No. 530001. During observation of the star in the order (we) on the tape of the chronograph there were registered the contact times $t_{9.9}, t_{10.1}, \ldots, t_{10.9}, \ldots, t_{11.1}, \ldots, t_{11.9}, t_{12.1}$. When the star was observed in the order (ew) the registered contacts were $t_{17.5}, t_{17.6}, \ldots, t_{17.9}, t_{18.1}, \ldots, t_{18.9}, t_{19.1}, \ldots, t_{19.6}$ (Figure 1).

We introduce the notation:

$$T' = \frac{1}{20}(t_{9.9} + t_{10.1} + \ldots + t_{10.9} + t_{11.1} + \ldots + t_{11.9} + t_{12.1})$$

and

$$T'' = \frac{1}{20}(t_{17.5} + \ldots + t_{17.9} + t_{18.1} + \ldots + t_{18.9} + t_{19.1} + \ldots + t_{19.6}).$$

Then for each star we get, depending on the order of observation, the average times T'_w and T'_e or T''_e and T''_w. The average distances of the chosen contacts from the n. c. l. were determined from the formulas:

$$f_{we} = \frac{1}{2}(T'_e - T'_w)\cos\delta,$$

$$f_{ew} = \frac{1}{2}(T''_w - T''_e)\cos\delta.$$

We note that the passage times were also calculated from the same initial data:

$$T_{we} = \frac{1}{2}(T'_e + T'_w),$$

$$T_{ew} = \frac{1}{2}(T''_w + T''_e).$$

The calculation of the errors in the half differences can obviously also throw light upon the errors in the half sums, i. e., on the passage times to be determined.

z	δ	f_{ew}	z	δ	f_{we}
61°	63°_{H}	$22^s.642$	10°	46°	$18^s.762$
12	68	22.621	5	51	18.762
28	28	22.651	27	29	18.766
1	57	22.585	27	29	18.776
35	21	22.648	55	69_{H}	18.624
39	17	22.664	17	73	18.748
17°	39°	22.589	25°	31°	18.767
	Average	22.657		Average	18.744

We give the results of the calculation of collimation of f_{ew} and f_{we} from observations of 14 stars on August 16, 1955, based on one clock error. In the considered example the observation in the lower culmination is marked with the letter H. The considerable deviations in the individual values of f_{ew} and f_{we} can only be explained by the variation of collimation or by the inequality of the pivots, because in the above-mentioned article it is shown that it is impossible to explain these deviations by variations of declination or of azimuth.

It is also impossible to explain them by the inequality of the pivots, which in this instrument does not exceed $0^s.01$.

In order to show the variation of collimation with time we carried out calculations of f_{ew} and f_{we} based on the observations of two clock errors in one night. Then we obtained the difference of the absolute mean values of collimation:

$$\Delta f_{ew} = (f_{ew})_2 - (f_{ew})_1 \text{ and } \Delta f_{we} = (f_{we})_2 - (f_{we})_1.$$

The results of these calculations are given in Table 1. This table also shows the date of observation, the temperature of the air, the mean values of f_{ew} and f_{we} and their sums $f_{ew} + f_{we}$. As the observation of the two clock errors was carried out without interruption, the differences Δf_{ew} and Δf_{we} show how the collimation changed on the average in one hour of observation. Table 1 shows that there are no seasonal variations of collimation for the instrument used. The change of collimation from the beginning to the end of observation is conclusive. The differences Δf_{ew} are negative and Δf_{we} positive.

TABLE 1

Collimation Values for APM-1 No. 530001

1955 dates day · month	t°	f_{ew}	Δf_{ew}	f_{we}	Δf_{we}	$f_{ew} + f_{we}$
20.4	+ 2.8	22.552	$-0^{s}027$	$18^{s}826$	$+0^{s}068$	$41^{s}378$
20.4	+ 2.1	22.525		18.894		41.419
4.5	+ 7.6	22.527	−0.031	18.836	+0.019	41.363
4.5	+ 6.4	22.496		18.855		41.351
20.6	+13.4	22.618	−0.077	18.810	+0.036	41.428
20.6	+12.4	22.541		18.846		41.387
1.8	+22.3	22.687	−0.059	18.706	+0.061	41.393
1.8	+21.0	22.628		18.767		41.395
8.8	+18.7	22.675	−0.042	18.739	+0.015	41.414
8.8	+17.5	22.633		18.754		41.387
15.8	+19.3	22.699	−0.056	18.719	+0.057	41.418
15.8	+18.0	22.643		18.776		41.419
16.8	+22.0	22.657	−0.012	18.744	+0.039	41.401
16.8	+20.5	22.645		18.783		41.428
2.9	+18.1	22.693	−0.043	18.704	+0.048	41.397
2.9	+17.2	22.650		18.752		41.402
12.9	+18.6	22.688	−0.029	18.697	+0.056	41.385
12.9	+17.4	22.659		18.753		41.412
4.10	+10.0	22.576	+0.008	18.834	−0.002	41.410
4.10	+ 9.6	22.584		18.832		41.416
17.10	+ 5.6	22.634	−0.022	18.772	+0.017	41.406
17.10	+ 4.9	22.612		18.789		41.401
5.12	−18.6	22.620	−0.144	18.740	+0.164	41.380
5.12	−19.2	22.476		18.904		41.380
Average	—	—	−0.044	—	+0.048	41.399

It is interesting to note that on observing the same stars on different nights, Δf was different. Such observations were made on August 15 and 16, September 2 and 12, and October 4 and 17 (Table 1). The change in collimation was particularly marked on December 5, 1955 (Table 1). It must be admitted that during that day the temperature was about 0° C and by nightfall a sharp frost had begun. It is possible that the large differences of collimation in the observation of the first and second deviation can to a certain extent be explained by the drop of temperature. We are inclined to suppose, however, that purely mechanical displacements of objective, prism, and micrometer are of no less importance here. When not in use, the objective of our transit instrument was at the bottom and during observation it was at the top. It is possible that transfer of the tube from the resting to the working position caused a slow change of the relative positions of the parts of the instrument and

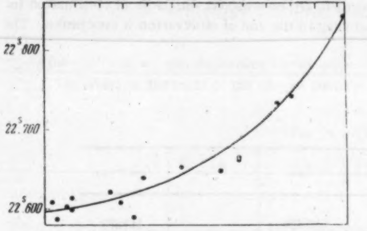

Fig. 2. Dependence of f_{ew} on the zenith distance for the observations on August 16, 1955.

z	f_{ew}	v
0°	$22^{s}.597$	— 66
5	22.603	— 60
10	22.609	— 54
15	22.616	— 47
20	22.626	— 37
25	22.638	— 25
30	22.653	— 10
35	22.671	+ 8
40	22.692	+ 29
45	22.716	+ 53
50	22.746	+ 83
55	22.784	+121
	22.663	—

consequently the collimation also changed. If the variations in f_{ew} and f_{we} were only to be ascribed to a temperature decrease, then a considerable seasonal variation would be observed. The results of Table 1 show that the variation of collimation cannot be explained by a temperature influence alone. It is also quite impossible to explain the dependence of collimation on zenith distance as the result of thermal influences on the instrument.

In the present example of absolute collimation calculated from the observation of one error of the clock, the zenith distances are also indicated. If we plot $\underline{z}$ and f_{ew} on a graph, we notice that with increasing $\underline{z}$, f_{ew} also increases. In Figure 2, f_{ew} is shown as a function of $\underline{z}$, based on the observations of two clock errors on August 16, 1955. Similar graphs were drawn based on the data of twelve nights of observations. After drawing a smooth curve, the values of f_{ew} for zenith distances 5° apart were taken from the graphs. The collimation values taken from the graph (for our example) are given in the small table.

TABLE 2

Deviations from the Average Collimation for Different Zenith Distances (in thousandths of seconds)

z / dates, 1955	0°	5	10	15	20	25	30	35	40	45	50	55°
20.4	— 82	—70	—54	—40	—26	—13	+ 1	+17	+35	+55	+ 77	+104
4.5	— 84	—68	—52	— 37	—22	— 7	+ 8	+22	+37	+52	+ 68	+ 84
20.6	—105	—92	—77	—61	—45	—26	— 5	+19	+45	+76	+115	+158
1.8	— 78	—70	—60	—50	—37	—24	— 8	+11	+33	+60	+ 92	+132
8.8	— 82	—77	—72	—64	—53	—38	—19	+10	+46	+81	+115	+153
15.8	— 64	—61	—58	—50	—40	—27	—12	+ 6	+25	+55	+ 91	+135
16.8	— 66	—60	—54	—47	—37	—25	—10	+ 8	+29	+53	+ 83	+121
2.9	— 51	—50	—47	—46	—41	—34	—23	— 8	+13	+44	+ 88	+153
12.9	— 78	—70	—61	—52	—40	—28	—13	+ 6	+28	+57	+ 94	+151
4.10	— 55	—49	—42	—35	—28	—21	—12	+ 1	+20	+44	+ 73	+102
17.10	— 67	—59	—50	—39	—29	—19	— 8	+ 8	+28	+53	+ 77	+101
5.12	— 66	—54	—43	—32	—19	— 8	+ 4	+17	+30	+44	+ 56	+ 70
average	— 73	—65	—56	—46	—35	—22	— 8	+10	+31	+56	+ 86	+122

The collimation difference between the zenith and $z = 55°$ is equal here to $0^s.187$. Such collimation variations were also observed on other nights.

Table 2 shows for each star the differences between f_{ew} and the mean value of f_{ew}, which characterize the collimation variation with the zenith distance.

As is seen from Table 2, f_{ew} increases with increasing zenith distances in all cases. When we plot $\underline{z}$ and f_{we} on a graph, we notice a reciprocal relation between them, i. e., with increasing zenith distance f_{we} decreases. The collimation at $z = 0°$ differs from that at $z = 55°$ by an average of $0^s.2$. This collimation variation is most serious, because the observation of stars is carried out at different zenith distances and thus evidently with continuously varying collimation.

Unfortunately, up to now it has been impossible to determine the collimation variation during the observation of a single star. However, there cannot by any doubt that the assumption that collimation for each star can be eliminated by reversal of the instrument is seriously in error.

A further increase in accuracy of astronomical observations with transit instruments is impossible without careful study of and allowance for the collimation errors.

P. K. Shternberg State Astronomical Institute

Received July 12, 1956

INCREASE OF THE SENSITIVITY OF SOME PHOTOGRAPHIC EMULSIONS BY MEANS OF PRELIMINARY SHORT EXPOSURE

T. M. Muliarchik and K. I. Petrova

The increase of sensitivity of three types of photographic emulsions by brief preliminary illumination before the main exposure is described.

Films of high sensitivity are usually intended for very short exposures ($t \approx 1/100^s$). They are inefficient for photographing faint objects through long exposures because of strong deviations from the reciprocity law.

A well known method for reducing these deviations consists in two successive exposures, the first of which is a short bright flash of light and the second a long exposure of low intensity [1, 2]. When registering brief processes as well as photographing very weak luminosities, such a double exposure gives an increase in sensitivity. In the first case, the emulsion must be submitted to a weak, long, uniform exposure after the main one [3]; in the second case it must be exposed uniformly during a short time to an intensive source before the main exposure [4]. Such a preliminary exposure can have a strong effect in photographing spectra of the night sky.

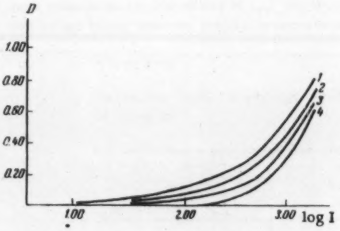

Fig. 1. Characteristic curves of film "type 10" for different preliminary-exposure densities.
1) D = 0.40; 2) D = 0.25; 3) D = 0.10;
4) D = 0.00 (nonpre-exposed part).

A selection of the most favorable pre-exposure conditions was carried out for two types of films — PF-3 and "type 10," and also for Isopan ISS plates. The analysis of the films was done with the sensitometer FSP-4 GOI (with ordinary light). The films were developed in a concentrated Chibisov developer, recommended for PF-3 (8 min at 20° C). The preliminary exposure density varied from 0.03 to 0.40, the time from $1/10000^s$ to 4^s.

The nonpre-exposed parts of the film and those pre-exposed to a certain density were each time simultaneously exposed under the same conditions. Finally the characteristic curves were constructed; the transmission logarithms of the step wedge of the sensitometer were plotted along the x-axis and the density of the corresponding steps of the wedge minus the fog density were plotted along the y-axis. For a pre-exposed part of a film the fog density is the sum of development fog and the density of pre-exposure.

The characteristic curves of the parts with different pre-exposure densities are represented in Figures 1 and 2 for "type 10" and PF-3, respectively. In all cases, the preliminary exposure was $1/100^s$ and the principal exposure was 1 hour.

The analysis of the influence of pre-exposure on the sensitivity of Isopan ISS plates was carried out with spectograph ISP-51, because the selectivity of the sensitizing mechanism was also examined. The plates were

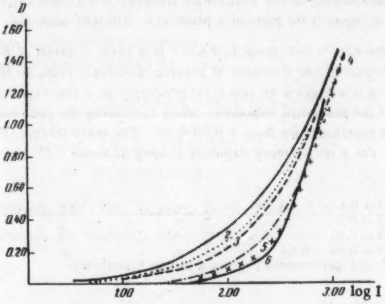

Fig. 2. Characteristic curves of film PF-3 at different pre-exposure densities. 1) D = 0.40; 2) D = 0.30; 3) D = 0.15; 4) D = 0.08; 5) D = 0.03; 6) D = 0.00 (nonpre-exposed part).

pre-exposed to ordinary light. The preliminary-exposure density varied from 0.05 to 0.40. One half of a plate was pre-exposed and after that, spectra of an incandescent lamp were projected on both parts. In order to extend the range of densities, several spectrograms for different slit widths were projected on each half of the plate. A common characteristic curve was obtained by shifting the individual curves along the x-axis.

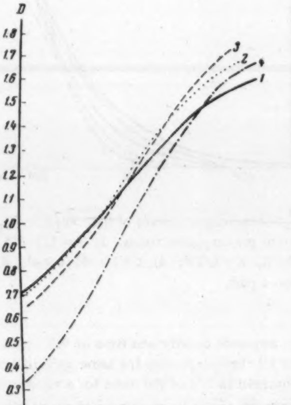

Fig. 3. Characteristic curves of Isopan ISS at different preliminary exposure densities. 1) D = 0.25; 2) D = 0.10; 3) D = 0.05; 4) D = 0.00 (nonpre-exposed part).

The preliminary exposures were of $1/50^s$ and $1/100^s$ while the principal exposure lasted 10^s. The spectra were measured with a photometer in the wavelength interval λ = 4300-6300 A. The characteristic curves with λ = 6300 A are presented in Figure 3 for parts of a plate with different densities.

The common feature of the curves in Figures 1, 2 and 3 is a lower contrast of the sensitized layer and an extension of the underexposed region in the direction of lower intensities. Thus, an intensity which on a nonpre-exposed film gives no blackening at all gives an appreciable density on a pre-exposed layer, the magnitude of which depends on the density of the preceding exposure. When increasing the preliminary exposure density, the sensitization increases. It has a maximum for D_{pe} = 0.30-0.40. The sensitization is different for different densities of the main exposure. For a preliminary exposure density of about 0.30, the gain in sensitivity is:

For $D = 0.4-0.5$ in PF-3 — 2 times, in ISS — 1.5 times
» $D = 0.3$ » 3 " 2 »
» $D = 0.1$ » 5 " 3 »
» $D = 0.05-0.03$ » 8 " 10
» $D < 1$ pre-exposure gives a loss in sensitivity

Figure 4 shows the relation between the sensitization and the length of preliminary exposure. The preliminary exposure density of 0.30 was obtained by exposures of $1/1000^s$, $1/100^s$, $1/25^s$, 1^s, 4^s. Pre-exposure of $1/100^s$ and $1/25^s$ produces the greatest effect. An exposure after the main one gives no increase in sensitivity of the emulsions.

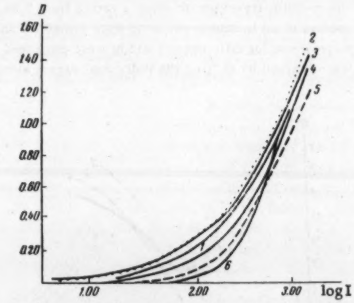

Fig. 4. Characteristic curves of film PF-3 for D_{pe} = 0.40 for different pre-exposure times. 1) t = $1/1000^s$; 2) t = $1/100^s$; 3) t = $1/25^s$; 4) t = 1^s; 5) t = 4^s; 6) the nonpre-exposed part.

For PF-3 film with preliminary exposure density and time of 0.1 and $1/100^s$ respectively, the main exposure varied from 1^s to 4 hours. For all these exposures the same gain in sensitivity was obtained. On a pre-exposed plate a density of 0.2 was obtained in 1/4 of the time for a nonpre-exposed one. Measurements of the spectrograms on ISS plates did not show the effect to be dependent on wavelength.

Thus, if the considered emulsions are used for emission spectra of the night sky and other weak objects, a preliminary exposure of the film to a density of 0.2-0.3 at an exposure of $1/100^s$ allows a shortening of the exposure by several times.

The pre-exposure must be carried out immediately before the principal exposure, because the increased sensitivity is retained for only a short time.

Institute of Atmospheric Physics, Academy of Sciences,
USSR, P. K. Shternberg State Astronomical Institute

Received August 25, 1956

LITERATURE CITED

[1] Webb and Evans, J. Opt. Soc. Am. 28, 431 (1938).

[2] Berg and Burton, Phot. J. 86, 13, 2, 62 (1946).

[3] Berg, Sci. Ind. Phot. 18, 124 (1947).

[4] Broun, Progr. Scientific Photography. Vol. 1 (1951) [in Russian].

THE EARLY THERMAL HISTORY OF THE EARTH

V. G. Fesenkov

Phenomena directly connected with the thermal history of the earth are considered: the flow of granite from magma, the origin and formation of continents, the dissipation of the primary atmosphere, the formation of oceans, the possible significance of underwater ridges, the motion of the poles on the earth's surface, the internal structure of the earth from seismological data and the earth's magnetism, the distribution of radioactive elements in the interior of the earth and the possible production of heat by radioactive processes according to data of different authors. Taking these phenomena into account, an attempt is made to represent the initial state of the earth and its formation from a primary nebula of gas and dust, surrounding the sun.

Considerations are given in favor of the loss by the earth of a considerable part of its mass, mainly light elements, during its formation.

Reasons for the differences between Mars and the earth are given.

The problem of the early thermal history of the earth is shared equally by geology and astronomy and must be studied by both sciences. Geologists must provide facts concerning the observed characteristics and structure of the crust and the inner shells of the earth. Astronomers can relate the earth's thermal history to the existing data on the formation of stars and planets of our galaxy and also, in a measure, draw comparisons between the earth and other planets of the solar system. It is essential in these studies to adhere strictly to facts and avoid preconceived opinions, which are especially dangerous in cosmogony, which has not yet freed itself from a priori hypotheses. First of all, then, the geological and astronomical facts pertinent to the problem must be summarized.

1. The existence of great granitic masses of which the continents are composed makes it necessary to discuss the origin of granite. Granite consists essentially of three minerals [1], quartz, feldspar and mica. It is the product of slow cooling of magma which can occur only at some depth; it has a characteristic texture and a low melting temperature. Many attempts have been made to duplicate this texture in laboratories, and different hypotheses have been advanced to explain its origin [2]. Without going into details of these investigations, it will suffice to say that apparently the only way in which granite can form is by crystallization from hot magma. Other theories of the origin of granite, by metasomatism of sediments or by migration of ions through solid rocks, are unacceptable, according to O. T. Tuttle. Granite, which constitutes the bulk of the outer shell of the earth—sial—must have formed at a temperature not lower than 1000° C. This requires explanation in the discussion of the early thermal history and evolution of our planet. As noted by P. N. Kropotkin [3], according to D. S. Beliankin granite may form as the result of a prolonged heating of siliceous rocks which, remaining in the solid state, exude a combination of oxides corresponding to the fraction with the lowest possible melting temperature. For this reason the melting temperature of granite is about 600° C lower than that of the ultrabasic rocks. Sometimes the opinion is expressed that granite may have formed from sediments fused at great pressures. Such is the view, for example, of D. I. Shcherbakov [4]. V. A. Obruchev ([1], p. 196) holds, on the

contrary, that this is the mode of origin of granitic gneisses. In any case, there is no doubt that granite is the product of slow crystallization of melts.

Since granite is the main material of continents, the problem of its genesis is intimately associated with that of the origin and gradual differentiation of the continental masses.

2. The origin of the continents must also be discussed in connection with the general problem of the origin of the earth.

It has long been known that continental rocks differ in composition from those underlying ocean basins. The former are mainly granites, the latter are the denser basalts. The boundary surface between the crystalline basalts and the underlying, still denser noncrystalline material, the so-called Mohorovicic discontinuity lying at the base of the crust, is much farther from the surface under the continents than beneath the oceans. According to E. Bullard [5], this discontinuity is due to change in composition of the material and not to a polymorphic transition. This is known from experiments in which the temperatures and pressures corresponding to the actual depth of this material were duplicated.

It is clear that the sialic continents, being lighter, must be isostatically elevated relative to the general surface of the earth. Thus, the heavier basaltic foundation must necessarily be occupied by oceans.

The origin of the continents is by no means clear as yet, but some clues to it may be obtained from a study of intermediate cases, namely, island chains. The rocks of these archipelagoes are mainly granitic but they are underlain by relatively thin layers of basalts covered with sediments. As noted by P. N. Kropotkin [3], the view is widespread that in the irreversible evolution of the earth's crust a gradual growth of the relatively stable continental platforms has taken place at the expense of geosynclines. In the more active geosynclinal belts, intensive folding develops, and igneous activity brings magmas close to the surface. These processes are active mainly near the platform borders and lead to the gradual growth of continental masses by the addition of newly formed volcanic island chains. Of course, some parts of the continents may undergo local transgressions, but the general trend of the processes remains unchanged — the granitic continental mass increases in volume and area. The irreversibility of the evolution of the crust was very convincingly demonstrated by V. A. Magnitski [6] (see also [7]).

Similarly, M. Kay [8] has pointed out that the island chains are very probably embryonic continents and illustrated this by the example of North America. Near the eastern and western margins of this continent there are volcanic chains which at one time were archipelagoes but later were joined to the continental block. In the regions of present day California and Florida there existed in the past only separate volcanic islands.

It is believed by some that in the beginning the entire surface of the earth was covered by the ocean, and that from it gradually rose separate islands which grew by addition of sialic material slowly formed in the depths of the earth and then brought to the surface.

The rise of magma from the depths, the development of platforms and the gradual diminution of geosynclinal belts represent an irreversible slowing down process. It is characterized by gradual diminution of fissure volcanism as well as volcanism in general throughout the history of the earth. In the past, as A. P. Pavlov [9] so vividly described, great lava fields were formed, over millions of square kilometers; they can be seen now on the Yenisei and Tunguska rivers, in Greenland, at Spitzbergen, in India, in parts of North America and elsewhere on the globe.

This process was accompanied by emanation of water vapor which contributed to the filling of the oceans.

The development of continents must have been most intensive in the earliest epochs of the earth's history. That enormous masses of silicate melt rising from the depths of the earth contributed to it is evident, but the dynamics of the process is by no means clear.

V. V. Belousov has pointed out that formation of granite occurred in the main in pre-geologic time [10]. In his opinion, the process of continental growth as described above is acceptable for the Pacific Ocean but not for the Atlantic and Indian oceans, which are much more recent features [11].

3. In order to discuss the earliest history of the earth it is necessary to consider the problems of formation of the atmosphere and the ocean basins. There is no doubt that the present atmosphere is secondary.

It is known that oxygen was produced by green plants and to a large extent by oceanic plankton; nitrogen is also partly due to the activity of plants but mainly it came from the depths of the planet from which, to a considerable extent, it comes now.

Only a few traces remain of the primitive atmosphere of the earth: they are the negligible quantities of inert gases such as neon, krypton and xenon, which do not enter into combinations with other elements. It is believed that helium, which is mainly a product of radioactive decay, can not remain near the earth's surface but is dissipated into inter-planetary space because of the high temperature of the layer of dissipation at the height of 400-500 km.

A question may be asked: What process was responsible for the dissipation of the earth's primitive atmosphere which affected even such heavy gases as krypton and xenon? It is evident that if the gases were dissipated separately in accordance with their atomic weights, the heavy ones, very abundant in the universe, would have remained on the earth in their original concentration. It must be assumed, then, that the process of dissipation was determined essentially by such a light gas as hydrogen, the main constituent of the primitive atmosphere. Escaping into space, hydrogen carried with it all the minor constituents; it was responsible for the dissipation of the primitive atmosphere, and this process did not require high temperatures. The escape of the heavier elements, for example of calcium, aided by hydrogen is observed on the sun.

It may be supposed that water vapor, if it existed in any quantity in the primitive atmosphere, also escaped during this process of dissipation.

The process of dissipation was not restricted to the primitive atmosphere. It began much earlier during the time of formation of the earth. This concept is not based on a hypothesis. The main fact which speaks convincingly of the loss of a considerable mass by our planet, mainly in the form of light and uncondensed elements, is the existence of the moon, as will be considered in greater detail later.

Where did the water of the modern oceans come from? It must have been derived from the depths of the earth during the period of its cooling.

Every volcanic eruption is accompanied by great quantities of water vapor which immediately falls in torrential rains. During the extremely intensive magmatic activity which accompanied the formation of continental masses, the surface of the earth received vast quantities of water which gradually formed the oceans. Graton [12] believes that the content of gases in extrusive magma may be as high as 10% by weight, and Friedlander [13] thinks that it may be much higher.

Goranson has shown [14] that granitic magma under high pressure corresponding to the depth of 10-20 km may contain great amounts of water in solution, up to 10% by weight. With decrease of pressure, the capacity of magma to hold water in solution decreases rapidly and it separates as water vapor. Separation of water vapor from cooling lava is a familar phenomenon. Therefore, every rise of sialic material to the surface led inevitably to separation of vast amounts of water, which remained on the surface of the earth and filled the spaces between the relatively high continental masses.

The average depth of the ocean uniformly spread over the entire surface of the earth would be 4-5 km, and this mass of water must have been derived from the outer layer or crust of the earth approximately 20 to 30 km thick.

It is doubtful that this could have taken place on an initially cold earth only gradually warmed by radiogenic heat.

The presence of the oceans as we now know them served, therefore, as a strong argument for an initially hot earth cooled by intensive convection currents.

According to Graton, even modern volcanism points to the existence of a primitive volcanic source related to tectonic displacements and lying at a depth much greater than 100 km.

It is interesting that volcanoes, both ancient and modern, are not isolated from each other but occur in groups covering large areas. In each group, individual volcanoes are alike in type and petrography, although the country rocks within the limits of the group may vary. Graton considers this to be evidence in favor of the existence of substratum of uniform composition which circles the globe (see p. 93 [12]).

4. Very interesting features which it may be supposed have a direct relation to the thermal history of the earth are: deep trenches, valleys and mountain ranges which are distributed in regular sequences on ocean floors and apparently remain unchanged for long periods of time.

Besides these mountain ranges, there are deep canyons on the ocean floors, sometimes having numerous tributaries, such, for example, as the canyon which starts near Greenland and extends south along the bottom of the Atlantic ocean to the 38th parallel.

The question of their origin was long ago posed by Academician L. D. Berg. At present it has been proved on the basis of new data that such canyons are formed by gigantic turbidity currents initiated by slides on the margins of continental platforms during certain times in the earth's history, as for instance at the end of the Ice ages, resulting from rapid melting of ice, or during strong earthquakes. Under normal conditions the velocity of submarine currents diminishes with depth and becomes negligible [15].

The enrichment of oceans with different salts necessary to prepare a medium for the inception of life could have occurred only with the participation of the continents, whether we consider the great water cycle or nature of the submarine currents.

More than ten deep trenches have been mapped in the Pacific and these may be intimately related to the origin of the continents. The general tectonic character of the oceans was described by Belousov [16]. Academician D. I. Shcherbakov [17] has pointed out that there are many submarine mountain chains in the region of the Antarctic continent forming a complex system which ties it to South America and partly to Australia. Huge mountain chains are separated by deep troughs, such as the South Hawaiian depression with its depth of 8264 m, for example.

Many of these trenches and mountain chains exceed in size those existing on the earth's surface. There are, for example, four great escarpments (Mendocino, Murray, Clarion and Clipperton) which start near the shores of North and Central America and trend westward at a considerable angle to the meridians, parallel to and nearly equidistant from each other [18]. These and other fracture zones characterized by sharply defined mountain ranges must have been formed as the result of tremendous dislocations of the earth's crust.

According to G. Menard [19], these structures could have been formed by huge convection currents immediately below the crust. It is assumed that hot materials from the depths of the earth rise and the relatively cold materials just below the crust sink. This phenomenon, Menard points out, has never been observed directly, for it could have taken place only during the earliest history of the earth, but it is the most likely explanation for the existence of these structures. The drag exerted by the convection currents against the lower surface of the earth's crust could have produced parallel fracture zones. It may be concluded that in the remote past the outer shells of the earth were plastic, a condition which no longer exists.

5. The distribution of radioactive elements within the earth has an important bearing on its history.

It is well known that the main radioactive elements, uranium and thorium, are enriched in the uppermost layers of the earth some tens of kilometers in thickness and that the abundance of these elements is greater in the younger rocks. V. G. Khlopin [20] presented convincing data on this subject. Meteorites, the analogues of the deeper parts of the earth, are only weakly radioactive.

It is natural to suppose that the primary material of the earth and other planets had a uniform distribution of the heavy radioactive elements simply because it was in a dispersed state. This raises the question, what process resulted in the present peculiar distribution of these elements.

There is no doubt that if the entire earth had the same concentration of radioactive elements as its surface layers it would never have solidified, and yet seismological data show convincingly that the earth is solid except for the inner core.

The cause of the selective distribution of radioactive elements, according to Goldschmidt [21], is the great size of the uranium and thorium ions, which does not allow them to enter into the silicate lattices during the process of crystallization of magma.

Therefore, during the initial stages of solidification of the deeper parts of the earth these elements were gradually forced upward, and enriched its upper layers. In this connection, it is interesting to note that, in

1928, Pigott [22] had pointed out that in granite the deposition of radium takes place on the surfaces of mineral grains and not within them. Holmes [23] has postulated that the volatile matter emanating from the depths of the earth carries radioactive elements with it.

This leads to the conclusion that the radioactive elements were uniformly distributed in the primitive earth while it was in a molten state. As it gradually solidified, the radioactive atoms migrated toward the surface and became concentrated in the crust. If it is assumed that the earth was cold in the beginning and never reached the molten state, the cause of the present distribution of radioactive elements cannot be explained. Their distribution is easily understood, however, if it is assumed that in the beginning the earth was molten and was cooled by intensive convection currents causing the radioactive elements to migrate upwards as it gradually passed into the solid state.

A detailed discussion of migration of radioactive elements was given by V. V. Belousov [24].

6. Perhaps the problem of the wandering of the poles should also be mentioned. Evidently no noticeable shifting of the poles occurred during the Tertiary. This was emphasized in various papers during the symposium at Cambridge devoted to this problem, and especially by Professor Dershem (University of California) on the basis of his study of fossil marine fauna of different epochs and by Dr. Arnell (Cambridge) and S. Runcorn, on the basis of the paleomagnetic studies [25]. According to them, several reversals of polarity occurred during the Tertiary but no change in position of the poles.

It appears probable, however, that before the Tertiary the average position of the magnetic field changed slowly with time. Runcorn explains this by slow shifting of the earth's axis.

Other specialists believe that the evidence is still insufficient for them to accept the wandering of poles as a fact since in the past, as now, the Arctic Ocean was characterized by low temperatures. Corals did not live there and deposition of limestones was relatively slight.

If, however, wandering of the poles is accepted, it must have occurred in distant geologic times when large movements of masses within the earth were still possible. A theoretical analysis of this latter phenomenon was made by the astronomer of the Greenwich Observatory, Gold, but factual evidence is scant. Undoubtedly, shifting of masses within the earth and changes of relief on its surface must affect the position of the earth's axis.

7. Of great importance to the discussion of the early history of the earth is the knowledge of its internal structure obtained by seismological methods in the past few decades. The basic facts are that the earth is solid throughout its entire mantle, from the core to the surface, and that the mantle consists of at least seven shells with densities changing abruptly from one to the next. These shells, according to Bullen, differ in chemical composition and can be only partially explained by polymorphic transitions. The outer shell is composed of olivine. In the deeper shells change takes place from one modification of olivine to another. Still deeper, heavier compounds dominate (Si — Mg — Fe), and beneath that lie shells of a variable composition. Both longitudinal and transverse waves pass freely through the mantle. Longitudinal waves do not pass through the core, and this is the main proof of its liquid state. It has been known since 1936 that the core itself is not homogeneous but consists of at least two concentric parts. The inner and denser part may be solid. According to Bullen [26] (see also [27]), it consists of iron-nickel, possibly with an admixture of denser materials.

As long ago as 1939, V. K. Lodochnikov suggested that with sufficiently great pressure, materials of the earth must pass into a metallic state stable in that environment. A more detailed investigation of this was made by Ramsey [29, 30]. The metallic state is marked by high electric conductivity because of weak bonds between atoms and electrons. This is accompanied by a sharp increase in density (for example, density of hydrogen in the metallic state increases sevenfold, and helium becomes 10-20 times denser). The metallic state involves destruction of electronic shells and of saturated molecular bands, and is characterized by a greatly increased electrical conductivity. There is no need, therefore, to think that the core is metallic in the usual sense, that it consists of a substance like an iron-nickel alloy which behaves like a metal under normal conditions of temperature and pressure. The change from the mantle to the core may simply be a change of state and not a chemical change, although either is possible. Since theoretical calculations related to the change into the metallic state are extremely difficult, although not impossible in the simpler cases, the final solution may be reached by comparing the earth with other planets. It is possible, according to Ramsey, to show that the observed mass, size and oblateness of Mars indicate that it is composed of the same materials as the earth if the difference between the two planets arising from a different distribution of densities with depth is taken into consideration.

On the other hand, Jeffreys has shown that the concept of similarity in chemical composition between the earth and other planets is invalid if the core of the earth has a composition substantially different from the rest of it. Admitting that all planets are composed of the same material modified by the existing pressures it is still possible to assume a considerable admixture of the heavy elements in the central parts of our planet. This is not based on observed facts but appears probable in view of the known differentiation of material in the mantle and is suggested also by the striking difference between the metallic and stony meteorites which are undoubtedly remnants of a planet or of several planets which once existed in the solar system.

8. The magnetic field of the earth provides evidence for judging of the properties of its interior. That the secular variations of the magnetic field of the earth originate entirely within the earth has already been shown by Gauss. According to the modern view the magnetic field of the earth is generated by electric currents within its liquid core [31]. This theory was worked out in 1939 by El'sasser [32]. The existence of the magnetic field shows that the core is liquid and not merely plastic, since varied and intensive movements occur within it. The change in polarity repeatedly observed in "fossil" magnetism of the Tertiary suggests that these currents may die out momentarily and then flow again but in the opposite direction.

All this points to the high temperature of the earth's interior, which must have remained practically unchanged since the beginning, partly because of low thermal conductivity of the earth's mantle and partly because of lack of exchange of material between the core and the mantle.

These are conclusions concerning the internal structure of the earth based on an analysis of existing data. It is possible to show that they are not compatible with the concept of an initially cold earth. On the contrary, very high temperatures must have existed on our planet from the very beginning.

Let us consider this in detail.

Four basic sources of heat existed during the evolution of the earth. One is the heat generated when a compacted body forms from a dispersed mass. If the process of agglomeration is rapid then not less than one-half of the kinetic energy developed from the potential energy during the process of compaction is heat.

The second source is radiogenic heat, which accumulates very slowly and whose effect becomes significant only over long periods of geologic time. When generated within the inner portions of the earth, nearly all of it is used up in heating, for little can escape upward. Thermal conduction becomes perceptible only at small depths below the surface. The effectiveness of this source of heat depends on the law of distribution of radioactive elements within the earth.

The third source of energy during formation of planets derives from chemical reactions. The importance of this source during formation of cosmic bodies has been repeatedly emphasized by Urey [33, 34].

To return our planet to its initial dispersed state, it would be necessary not only to overcome the gravitational forces holding it together but also to break chemical bonds of various complex molecules in order to obtain simple molecules and atoms of the protoplanetary cloud. The energy required for this would exactly equal the energy expended during the formation of the earth.

The birth of our planet involved not merely the aggregation of the originally dispersed matter but also formation of very complex molecules.

Finally, the gravitational differentiation of materials of the earth must have had a certain effect on its heat budget.

It can be easily shown that if gravitational condensation were the only process involved in the formation of the earth, even then sufficient amounts of energy would have been released to melt the entire globe. In a heated mass convectional currents would be readily started and lead to a rather rapid cooling. During this process radioactive elements would be carried towards the surface and become concentrated in a relatively thin layer. Thus, radioactive materials migrated toward the surface before the effects of radiogenic heat became perceptible in the interior and their influence on the heat budget of the earth as a whole became negligible. Concentration of radioactive elements in a thin surface layer affects the geothermal gradient and in case of unusual local concentration may possibly result in local fusion. Assuming the average geothermal gradient to be 30° C per kilometer, we find that the maximum temperature at the base of this layer cannot exceed 1000° C.

Similar temperatures must obtain throughout the earth, if it is assumed to be without any other sources of heat. With the tremendous pressure within the earth, such temperatures obviously can have but little effect.

In the past, when the amounts of radioactive materials were much greater, the geothermal gradient was much steeper and the temperature at the base of the radioactive layer was higher but still inadequate to cause melting of rocks at the base of the mantle where pressures of 1.3 million atmospheres exist.

Therefore, the exact way in which the earth was formed from a dispersed cloud becomes of primary importance in tracing its thermal history. This brings us to a discussion of the mechanism of formation of the planets.

According to the well-known cosmogonic theory of O. Iu. Schmidt [35], the earth and the other planets were formed from a cloud of meteoric dust by gradual agglomeration and welding together of individual particles colliding with one another. The original part of Schmidt's concept is in his postulate of capture of such a dust cloud by the Sun. This capture of a rather massive cloud has not been theoretically discussed and appears to be quite impossible. Moreover, such a mechanism is unnecessary, since the planets were formed not from a cloud consisting of dust alone but from an extensive and rather dense mass of gas in which dust particles constituted a minor admixture. There is no need to cite arguments for this concept, since no one is likely to deny it now.

Essentially the process of planet formation differs from that of the formation of stars only in scale, just as planets differ from stars merely in having much smaller masses.

There exist several calculations of the rate of heating of the primitive earth based on the postulate of its origin from a cloud of dust by accretion. Such a calculation in strict conformity to Schmidt's cosmogonic hypothesis was made by V. S. Safronov [36].

The rate of growth of an embryo planet depends on the frequency of its encounters with particles of dust and on the density of the cloud. Safronov calculated the rate of growth of the planet and as a corollary the rate of heating of its surface. The temperature of the surface for every given moment was determined by considering the balance between the heat generated by the fall of particles and the radiation of heat into space according to Stefan's law.

Having made probable assumptions as to the speed of the dust particles, he found that the temperature of the primitive earth during its formation could not have exceeded 1000° abs. at the depth of approximately 1000-1500 km.

The time required for the growth of the earth to its present size under these conditions was given as 20 million years. This is a short period of time compared with the age of the solar system, but it should be noted that the duration of solar radiation in the process of purely gravitational contraction of the Sun from an infinitely large size is also approximately 20 million years, according to well-known calculations. The relatively huge mass of the sun falling upon itself from infinity was capable of maintaining radiation without aid from atomic reactions for only this extremely short period of time.

The arguments of Safronov are not very convincing, since energy is dissipated not only during the process of accretion but during an indefinite period of time after the fall of a particle on the planet's surface and this must lead to an incomparably greater lowering of temperature. Thus, the result obtained by Safronov must be regarded as a certain maximum value only.

The author of the meteoritic hypothesis, Schmidt, considered the process of accretion insufficient for heating the planet (see Four Lectures on the Origin of the Earth [37]). In general, it was believed that in the light of this hypothesis the earth must have been completely cold in the beginning. Such was the opinion of P. N. Kropotkin [3]; E. A. Liubimova, in her calculations of radioactive heating of the earth, based on Schmidt's views, assumed in her principal work [38] that the initial temperature of the earth was about 0° C.

It is clear that the mechanism of formation of the earth proposed by Schmidt does not lead to a noticeable increase in temperature. In any case, this mechanism could not have resulted in the melting of the whole earth or of a large part of it, including the core which is still liquid, and therefore intensive convection currents could not have been generated within the earth. Analogous calculations were made by T. Gold [5] on the assumption that the earth grew by accretion.

All these calculations show that even in this process of formation of planets a certain amount of heating will take place, reaching perhaps 1000°C, but this is quite insufficient to melt the earth and especially its inner core and to insure rapid cooling. Therefore, in order to explain the present high temperature of the earth and especially of its interior, it was necessary to turn to a second source of heat, radioactive decay. A number of works by E. A. Liubimova in the Institute of Geophysics, Academy of Sciences, USSR, and also by some foreign authors [39, 40], are devoted to this problem.

E. A. Liubimova, like Safronov, starts with Schmidt's hypothesis by assuming that the primitive earth was cold when heating by radiogenic heat began, and implies a uniform distribution of radioactive elements throughoug the earth. This is a highly artificial approach since it makes it necessary to assume that the content of radioactive elements in the earth is as low as in meteorites and lower than that observed even in the deepest terrestrial rocks. This concentration corresponds approximately to that which the earth would have if all radioactive substances were concentrated in a layer 200 km thick.

Liubimova's calculations show that the thermal field of the earth can be divided into two regions: the outer, where heat flows towards the surface and is partly radiated into space, and inner where temperature increases continuously because of generation and conservation of radiogenic heat. The depth of the layer separating these two fields is of the order of $6\sqrt{kt}$ (k is the thermal conductivity; t is the temperature) and increases slowly with t.

At present, this depth is about 1000 km ($k = 0.01$).

These calculations show also that in the interior the temperature continuously increases, while in the outer region it decreases. The maximum temperature which could be obtained in the course of the earth's evolution from radiogenic heating of the lower part of the mantle at the core boundary is about 4000° C. Cooling of the outer part of the earth causes a pronounced contraction to the depth of 700 km, and this gives rise to stresses and earthquakes. From this point of view, in spite of the postulated distribution of the sources of radioactive decay, the earth has always remained solid. However, occasional local concentrations of radioactive substances could have led to local melting, to pockets which served as sources of igneous activity.

A number of known facts concerning the internal structure of the earth make it necessary for Liubimova to abandon the concept of an initially cold earth and to seek new factors which would insure sufficient heating. For example, in order to explain the concentration of radioactive substances in a thin surface layer, it is necessary to assume that the entire earth was heated and brought to a plastic state so that convection causing migration of radioactive substances to the surface became possible. However, this convection must occur only after a complete fusion of the core, since the extreme slowness of radioactive heating and the rapid escape of these substances would not allow time for heating the inner portions of the earth to a sufficiently high temperature. Such artificial constructions must be introduced merely because Liubimova is committed to Schmidt's concept of formation of the earth by slow aggregation of dust particles, a process which does not provide the primitive planet with the necessary heat energy.

Actually, it must be assumed that the earth and other planets were formed not from a quasi-meteoric cloud captured by the sun in passing through a hypothetical nebulosity, but from a gas and dust medium surrounding the primordial sun and identical with it in chemical composition.

It is quite evident from rather elementary considerations that during formation of a star from a gas and dust medium only a certain part of the condensing cloud goes into the making of the star; the remainder and possibly the bulk of the medium must remain in a dispersed state because too great an angular momentum cannot be concentrated in one body without its becoming unstable. Only under exceptional conditions could the entire mass of a cloud be gathered into one compact body. As a rule, a new star will be surrounded mainly in its equatorial plane by a cloud of gas and dust with a rather high density and a relatively low temperature.

Because of gravitational instability this cloud provides more than enough material for secondary condensations or planets. The angular momentum which cannot be concentrated in one compact body must necessarily, during the advanced stages of condensation, be distributed among separate gravitational centers with their own orbital motions. Even then, much of the cloud must remain in a dispersed state. Illustrations of this are numerous double and multiple stars in our galaxy and also the earth-moon system. It is almost certain that the moon, during its formation, acquired a considerable part of the initial angular momentum and thus allowed the earth to complete its condensation. But if this point of view, based on a modern concept of the formation of stars, is accepted, an elementary calculation shows that the modern earth represents only a small fraction of its original condensation. Indeed, if the initial angular momentum and consequently the mass of the earth had not been lost to a large extent the combined mass of the earth and moon united into one body with the total angular momentum which now characterizes this system would have been unstable and the period of its rotation would have been at most 1.4 hours. In reality because of this loss it is about 4.2 hours.

It may be supposed that the process of formation of a planet, once it started, would go on rather rapidly. According to modern calculations the time of condensation of a star from a globule is comparable to time needed for a simple fall of an element of mass from its surface to the center. To initiate this process it is necessary only that through a preliminary condensation the density of individual heterogeneities which must inevitably form within the diffuse cloud satisfied the condition of tidal stability, which for the central parts of the solar system, for example, amounted to about 10^{-6} g/cm^3.

The temperature of planetary condensation can be estimated from the fact that about half of the potential energy is expended on heating it during the process of shrinking.

In the case of a homogeneous sphere, the potential energy is

$$U = \frac{3}{5}\frac{GM^2}{R}.$$

In condensation of a heterogeneous sphere the expenditure of energy increases. For instance, in case of heterogeneity satisfying Roche's law, which formerly was widely applied to the earth, namely:

$$\delta = \delta_0\left[1 - \beta\left(\frac{r}{R}\right)^2\right], \quad \beta = 0.38,$$

we have

$$U = 3.6 \cdot 10^{39} \text{ ergs}$$

In the case of heterogeneity satisfying the exponential law

$$\delta = \delta_0 e^{-\alpha(r/R)},$$

we obtain

$$U = f(\alpha)\frac{GM^2}{R},$$

where

$$f(\alpha) = \frac{9}{8}\,\frac{11 + 22\alpha + 14\alpha^2 + 4\alpha^3 + 5e^{2\alpha} - 16(\alpha + 1)e^{\alpha}}{\alpha^5},$$

and for $\frac{\delta_R}{\delta_0} = 0.15$ we have $$f(\alpha) = 3.7.$$

In the case of modern earth:

$$\frac{GM^2}{R} = 3.75 \cdot 10^{39};\ R = 6400\ \text{km};\ M = 6 \cdot 10^{27}\ \text{g}.$$

All these values of potential energy are of the same order of magnitude and lead to heating measurable in tens of thousands of degrees.

In the Roche case, $1.8 \cdot 10^{39}$ ergs or $1.5 \cdot 10^4$ cal/gram are available for heating.

It is quite evident, then, that primitive heat generated during the process of formation of the earth is of the greatest importance in its later history. Radioactive heating is very slow and before its effect could become significant all radioactive substances would have escaped toward the surface where their role is localized and does not affect the entire earth.

Jeffreys, in the last edition of his well-known book [41], writes, in complete accord with the above, that initially the earth was gaseous and could have radiated all of its energy in about 500 years.

The total energy of condensation of the earth to its present radius is estimated by him as 10,000 cal/g and the energy due to initial temperature as approximately 6000 cal/g. In the presence of convection currents, the earth could have cooled to a liquid state in only 5000 years after its birth. With further cooling, which took place from the outside inward, the total time of cooling and solidification with cessation of convection was much longer. Jeffreys assumes the following conditions in discussing the process of cooling ([41, p. 286]: the initial temperature of the surface is near melting point and gradually decreases to 300° K; the radioactive layer is very thin and radioactivity decreases rapidly with depth.

The formulation of the problem given above agrees well with the modern concepts of science.

It remains to consider a number of side issues of great interest. First, why does not our neighbor planet Mars have water bodies and is poor in water in general?

It may be asserted at present that Mars can easily hold water vapor without allowing its escape into space. As was shown by L. Spitzer [42], in order to allow the escape of water vapor, the temperature of the appropriate layer of the martian atmosphere should exceed 500° K. Actually the temperature of the martian atmosphere is much lower than this. On earth, the temperature of the high ionosphere increases considerably because of the presence of atmospheric oxygen produced by photosynthesis. On Mars, oxygen is absent and no such heating effect can take place. The loss of water vapor which comes to the surface of the planet from its depth or exists in its atmosphere must occur inevitably in the earliest epoch of its history when the entire mass of the planet together with its surface is still hot [43]. However, the presence of only local regions of fusion on the surface of a planet is sufficient to effect escape if they are large enough to heat the atmosphere above them to the zone of dissipation, which on the earth is 400-500 km above the surface. Therefore it became possible for Mars to retain water only when the formation of its crust was far advanced.

The very small amount of water existing on Mars at present is evidently the product of volcanic eruptions which occurred after the planet had almost completely solidified. The moon, with a mass almost a hundred times smaller than that of Mars, differs from it merely in that it has not only completely lost water vapor during the process of cooling, but continues to lose it, if (as some selenologists believe) it still issues from its interior.

The evidence presented above leads to the conclusion that because of the mechanism of its formation, the earth (from the very beginning) had considerable energy and therefore high temperature, and that the slowly accumulating and gradually diminishing radioactive heat was of secondary importance for the earth as a whole.

Received December 7, 1956

LITERATURE CITED

[1] V. A. Obruchev, Principles of Geology [in Russian] (Moscow, 1956),p. 119.

[2] O. Fr. Tuttle, "The origin of granite," Sci. American, April (1955).

[3] P. N. Kropotkin, Priroda 4,31 (1956).

[4] D. I. Shcherbakov, Commentary, Collected Works of Jules Verne 2 (Moscow, 1955).**

[5] L. Hawkes, "Evolution of the earth," Nature 4457 (1955).

[6] V. A. Magnitskii, Problems of Cosmogony 1, 15 (1952).

[7] *

[8] M. Kay, "The origin of continents," Sci. American, September (1955).

[9] A. P. Pavlov, Volcanoes, Earthquakes, Seas, Rivers (Moscow, 1948).**

[10] V. V. Belousov, Fundamental Problems of Geotectonics [in Russian] (Moscow, 1954) pp. 68 and 512.

[11] V. V. Belousov, Fundamental Problems of Geotectonics [in Russian (Moscow, 1954) p. 510 ff.

[12] L. C. Graton, Conjectures Regarding Volcanic Heat [Russian translation] (IIL, Moscow, 1949).

[13] J. Friedländer "Physics of Earth", Volcanology 35 (Wash.,1911).

[14] H. Jeffreys, The Earth (third edition, 1952).

[15] H. Stommel. Sci. American, January (1955).

[16] V. V. Belousov, Fundamental Problems of Geotectonics [in Russian] (Moscow, 1954) p. 509.

[17] D. I. Shcherbakov, Prioroda 44, 2 (1956).

[18] R. L. Fisher and R. Revelle, "The trenches of the Pacific," Sci. American, November (1955).

[19] H. W. Menard, Sci. American, July (1955).

[20] V. G. Khlopin, Izvest. Akad. Nauk SSSR, Otdel. Mat. i Est. Nauk (1937).

[21] V. M. Goldschmidt, Naturwiss. 18 (1930).

[22] G. S. Pigott, Am. J. Sci. 5, 17 (1929).

[23] A. Holmes, Geol. Mag (1915).

[24] V. V. Belousov, Izvest. Akad. Nauk SSSR Ser. Geogr. i Geof. 1942, No. 6 and 1943, No. 3.

[25] A. A. Daly and S. K. Runcorn, "Polar wandering," Nature, 4479 (1955).

[26] K. S. Bullen, "The interior of the Earth," Sci. American, September (1955).

[27] J. A. Jacobs, "Inside the Earth," J. Roy. Astron. Soc. Can. May-June (1955).

[28] V. I. Lodochnikov, Zapiski Vsesoiuz. Mineral.Obshchestva, Ser. 2, 1939, No. 2, 3.

[29] W. H. Ramsey, Astr. Contr. from the Univ. Manchester 18 (1955).

[30] W. H. Ramsey, Monthly Notices 108 (1948).

[31] A. T. Kalashnikov, "Problems of terrestrial magnetism," Vestnik Akad. Nauk SSSR 1 (1955).

[32] S. K. Runcorn, "The Earth's magnetism," Sci. American, September (1955).

[33] H. Urey, The Planets (New Haven, 1952).

---

*Missing in Original.

**In Russian.

[34] H. Urey, Report at the 7th Colloquium (Russian translation) (Liege, 1956).

[35] O. Iu. Schmidt, Four Lectures about the Origin of the Earth [in Russian], (Moscow, 1950).

[36] V. S. Safronov, Astron. J. (USSR) 31, 6 (1954).

[37] O. Iu. Schmidt, Four Lectures on the Origin of the Earth [in Russian], chapter 4, p. 75 (1950).

[38] E. A. Liubimova, Izvest. Akad. Nauk, Ser. Geofiz 2 (1952).

[39] E. A. Liubimova, Izvest. Akad. Nauk, Ser. Geofiz 5 (1955).

[40] E. A. Liubimova, Doklady Akad. Nauk 107, 1 (1956).

[41] H. Jeffreys, The Earth, third edition (Cambridge, 1952).

[42] L. Spitzer, Jr. and Kuiper, The Atmosphere of the Earth and Planets (1952).

[43] V. A. Firsoff, J. Brit. Astron. Assoc. January (1956).

THE PHYSICAL NATURE OF THE DIFFERENCES BETWEEN GEOMAGNETIC DISTURBANCES OF SUDDEN AND NONSUDDEN COMMENCEMENT

E. R. Mustel'

The differences between sudden-commencement geomagnetic storms and nonsudden-commencement storms are considered. The discussion shows that the ejection of condensations, leading to nonsudden-commencement storms is evidently caused by some shortlived nonstationary process of electromagnetic nature in the region of sunspots. This process is always accompanied by peculiar phenomena of radio emission with a shift of the level of radio emission in the corona. It can be accompanied by chromospheric flares, but does not necessarily cause them. Disturbances with nonsudden-commencement characteristics are caused by flocculi. This is confirmed by the fact that the passage of flocculi through the center of the sun's visible disk is accompanied by geomagnetic disturbances [7, 8].

The difference in the character of sudden-commencement and nonsudden-commencement disturbances must be connected with differences in physical properties of the streams and not with their geometrical properties. Condensations causing sudden-commencement disturbances have larger velocities than those causing nonsudden-commencement disturbances; this influences the rapidity of the variations of the Earth's magnetic field (Formula 1).

Besides, condensations, causing sudden-commencement disturbances and ejected from the region of sunspots, have their own considerable "frozen" magnetic field. This should also cause additional variations of the Earth's magnetic field. Considerations show that the field of the condensations, causing nonsudden-commencement disturbances should be much smaller than in the first case.

Geomagnetic disturbances constitute one of the main sources of information on the corpuscles emitted from the sun. In the first place, such disturbances are basically corpuscular in character (except for such phenomena as the "hooks" observed at the time of chromospheric flares). In the second place, we have at our disposal singularly uniform data collected over many decades and pertaining to geomagnetic disturbances.

It has been determined by now that geomagnetic disturbances can be divided into two groups. The first of these consists of disturbances of sudden commencement, the second — disturbances of nonsudden-commencement. It is known that the intense disturbances which usually accompany brilliant chromospheric disturbances are characterized by sudden-commencement. However, the weaker geomagnetic disturbances also include those of sudden commencement; thus, weak and medium-intensity disturbances are subdivided into nonsudden-commencement ones (these are in the majority) and sudden-commencement ones (the minority). It is very important to note that these two types of disturbances differ not only in the character of their commencements but also in many other characteristics. Considering at first only the weak and medium disturbances, we will list the basic differences between the two types.

a) They are entirely different as to the character of their numerical distrubution in the cycle of solar activity. The number of the sudden-commencement disturbances follows the progress of sunspots in the cycle quite closely [1]. On the other hand, the curve corresponding to the number of the nonsudden disturbances is displaced by some three to four years in respect to the curve corresponding to the number of sunspots (the disturbances occurring later than the sunspots), and is altogether different in character [1].

b) The nonsudden-commencement disturbances are characterized by their 27 day sequence, which is especially pronounced prior to the minimum of solar activity. On the other hand, the sudden-commencement disturbances only rarely occur in such short sequences; note especially the work [2].

c) The most intense of all the weak and medium disturbances are of the sudden-commencement type.

d) The two types of disturbances differ, as is indicated by their very name, in the nature of their beginning phase. The same can also be said of their terminal phase [1].

e) The duration of the nonsudden-commencement disturbances is sometimes very great and may extend up to 10 days, while the sudden-commencement disturbances last usually from 24 to 36 hours. About 10 per cent of the latter disturbances last over 36 hours, and some of them are shorter than 24 hours.

f) The disturbances differ also in the character of their progress; those with the nonsudden commencements exhibit weak oscillations, usually with different phases of hydrogen and deuterium oscillations. The sudden-commencement disturbances oscillate rapidly or very rapidly, and the character of the oscillations is reversed from time to time. Furthermore, the sudden-commencement ones* often exhibit this storm-time variation, while the nonsudden ones do not show this feature.

g) Streams comprising the nonsudden disturbances are characterized by high-intensity radio emission. The following facts corroborate this:

1) A number of nonsudden-commencement disturbances exhibit a sharply defined seasonal regularity, with spring and fall maxima [3, 4].

2) These maxima are most sharply defined in years of minimum solar activity.

3) The curve representing the number of nonsudden-commencement disturbances is displaced (in the direction of retardation) in respect to the curve representing the number of sunspots. This is indicative of the importance of the low solar latitudes and, at the same time, the radio emission of the streams. There are also a large number of other factors indicative of radio emission.

The characteristics listed above pertaining to weak disturbances of sudden commencement are even more emphatic in the case of intense disturbances of sudden commencement. Thus, sudden-commencement disturbances generally differ in many respects from those of nonsudden commencement.

Everything that has been said above is indicative of the fact that the solar origins of the streams comprising nonsudden and sudden-commencement disturbances are altogether different. The following facts are expecially important in arriving at this conclusion: 1) the completely different character of the distribution of the two disturbance types in the solar cycle; 2) the difference in the character of the 27-day sequence, and consequently, in the persistence of streams in the two types; 3) the difference in the character of radio emission of the streams; i. e., geomagnetic disturbances related to bright chromospheric flares are produced even when a flare occurs at a distance of up to 45° from the center of the disc. This means that in this case the band of corpuscles is very broad. We may assume (see below) that the same is also true for the streams comprising weak disturbances of sudden commencement. 4) The very character of the disturbances is entirely different.

In this article we propose to investigate thoroughly the problem dealing with the origin of disturbances of both types. Let us consider the sudden-commencement disturbances. It follows from the facts listed above that these must be closely related to the sunspots. It is obvious, however, that the sunspots themselves are not the source of the corresponding streams. Also, in the case of intense geomagnetic disturbances of sudden commencement the source of the corpuscles is normally in the bright chromospheric flares, which in themselves are very closely related to the spots [1]. Furthermore, observations of solar radio emissions indicate directly that the solar corona at the time of a flare exhibits an outward-from-the-flare movement of some disturbed agent which, from all the indications, represents a flow of atoms away from the flare. It is possible also to conclude, on the basis of the same radio-emission observations, that atoms in the investigated stream have a very complex velocity spectrum.

---

*We have discussed here the weak disturbances. The storm-time variation is nearly always present in the case of intense disturbances of sudden commencement.

All of this refers to the intense geomagnetic disturbances. The matter of the medium and weak disturbances of sudden commencement is not quite clear. It would be possible to assume that these, too, are produced by the flares but with less intensity. Thus, for example, Newton [5] claims (see, in particular, Figure 2c of his article) that those spot groups within which chromospheric flares are most commonly produced are geoactive. At the same time, a number of data indicate that sudden-commencement disturbances can be observed without any apparent connection to chromospheric flares. This allows us to made the following suppositions.

The latest investigations make it possible for us to conclude that the outflow of atoms from chromospheric flares is not conditioned by the occurrence of the bright-line emission of the flare spectrum, but, on the contrary, that both these phenomena (that is, the emission of atoms and the appearance of a bright flare on the disc, or at its edge) represent the results of the same extremely complex process which is, apparently, of an electromagnetic nonstationary character.* In fact, if the emission of atoms from the flares were conditioned by the bright radiation of the flare, then the only mechanism capable of producing such an emission would be that of the light pressure. But the author has pointed out in one of his recent articles [6] a number of reasons making the acceptance of this concept difficult. There are also other even more serious objections to it; study of the velocity spectrum of particles emitted from the bright flares (see above) indicates that light pressure can be of no consequence here, because the spectrum includes, besides the velocities on the order of a thousand kilometers per second, considerably higher velocities of the order of several thousand kilometers per second, not to mention the emission of particles with the energies of cosmic rays.

But if all this is true, it is quite possible that the indicated nonstationary electromagnetic process is not necessarily accompanied by the appearance of a bright chromospheric flare on the disc. In other words, it is quite possible that in a number of cases (or in a great many cases) the activity of this process does not appear in all the possible forms, which is to say that this process may be accompanied by the emission of corpuscles, but may not be accompanied by the appearance of a chromospheric flare. Opposite cases are also possible. Indeed, observations of chromospheric flares and of radio emission confirm the complicated nature of the whole picture. Further proof of the fact that in all the cases of corpuscle emissions we deal with the same specific nonstationary process (regardless of whether a bright chromospheric flare does or does not appear) is provided by the identical character of the intense and weak disturbances of sudden commencement. This fact is of great importance.

It follows that for the sources of the corpuscles producing the sudden commencement distrubances we must, apparently, accept those regions in which those originate specific nonstationary phenomena related to the intensification of radio emission and to the displacement of the level of this radio emission in the corona. These phenomena do not have to be accompanied by the bright chromospheric flares. We can only venture a guess that the indicated regions are very ephemeral in nature, and that they are related to the sunspots (see, for instance, the previously mentioned results obtained by Newton [5]).

We will discuss briefly the origin of corpuscles producing geomagnetic disturbances with a non-sudden commencement. The author believes that the corresponding solar regions coincide with the regions of the flocculi light with the rays of H, K and Ca^+. This is confirmed by the fact established by the author [7], that the passage of nearly every flocculus (or facula) through the visible center of solar disc is accompanied for a certain period of time by geomagnetic disturbances. Consideration of these matters is contained in author's other articles [8, 9] and we will not stop to review them here.

---

*In fact it is possible that the phenomenon is even more complicated; it is possible that it encompasses the processes involving impact waves and others.

We will pass now to the main question for this article, which is the consideration of the nature of differences between the disturbances of sudden and nonsudden commencements. It is obvious that these differences are related to the differences in some properties of the streams producing the two types of disturbances. Allen [10] believed that these differences are conditioned by the variations in the character of entry of solar corpuscles into the upper layers of the earth's atmosphere; this is to say that in the case of disturbances with a sudden commencement, we are dealing with a relatively short period of emission of the corpuscles (lasting only as long as the formerly indicated nonstationary process), after which the corpuscles advance in a broad front of a definite breadth (Figure 1, a). On the other hand, in the case of disturbances of nonsudden commencement, in which the regions of emission of the corpuscles are relatively stationary and last for months, the intrusion of the corpuscles into the earth's atmosphere takes place because the stream of the corpuscles is transposed in space in the direction of solar rotation, and after every 27 days "overtakes" the earth. Allen believes that in the first case (Figure 1, a) the foremost border (forward front) of a stream is sharply defined, and that this is the cause for the sudden commencement in corresponding disturbances. In the second case (Figure 1, b) the earth falls into the "side" portions of a stream where the border is less sharply defined; this fact leads to disturbances of nonsudden commencement.

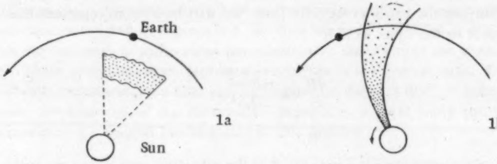

Fig. 1. Differences in the geometrical character of the streams producing geomagnetic disturbances of sudden and nonsudden commencement.

It is difficult however, to agree with Allen's view. In the first place, it follows from the very character of geomagnetic disturbances, whether their commencements are sudden or nonsudden, that the stream consists, in both cases, of separate condensations divided from one another by the zones of considerably lesser condensation of matter. Consequently, there is no reason for the existence of any definite border at all in the first case.

The second consideration rests on the previously mentioned fact that the two types of disturbances differ not only in their early and final phases, but also in the very character of their progress. Moreover, it follows from Figures 1a, and 1b that after the entry of the earth into a stream, its situation appears to be identical in both cases, because the following (after the commencement of the disturbance) progress of events is determined by the velocity of flight of the particles from the sun (velocity of atom movements in Figure 1, b along the tangent to the earth's orbit is insignificantly small in comparison with the velocity of the particles moving away from the sun).

The author of this article believes that the differences between the two investigated types of disturbances are conditioned by different physical characteristics of the corresponding streams. As such characteristics we must list primarily:

a) velocity $\underline{v}$ of the condensations and of the separate atoms in the streams,

b) concentration $\underline{n}$ of atoms in a stream and in the condensations,

c) magnetic field of the condensations in a stream.

These three basic characteristics will be considered below.

First let us pass to the question of the velocities of condensations in the streams. A number of data indicate that velocity $\underline{v}$ in the streams producing disturbances with sudden commencements is, on the average, greater (in many cases considerably greater) than velocities $\underline{v}$ in the streams producing disturbances with the nonsudden commencements.

As the disturbances with sudden commencement, we will investigate those disturbances which are related to the chromospheric flares. By basing our judgment on the retardation of geomagnetic disturbances (with respect to the appearance of the flare) we can estimate velocity $\underline{v}$ fairly accurately. On the average, this period of retardation is equal to 20 hours [11], which gives us the velocity of about 2000 km/sec. (although greater velocities can also be observed). On the other hand, velocity of the corpuscles in the streams producing disturbances with a nonsudden commencement is undoubtedly lower. In the same way, a study of disturbances produced by the passage of the flocculi over the center of the solar disc gives us the period of retardation equal to a few days, and sometimes up to ten days.*

Similar results, pertaining to the differences in the velocities of condensations in the streams of the two types, are obtained from a study of the shift of line H_α in the spectrum of the Aurora Borealis (see [12], p. 412).

At the same time the difference between the velocities of atoms and condensations in the streams is of a great significance in determining the character of geomagnetic disturbances. If we should inspect a simple model depicting the first phase of a geomagnetic disturbance, a model in which kinetic energy in the corpuscles of a stream is expended on changing the earth's magnetic field, we will be able to represent the change in the intensity of this field with respect to the time, as [13]:

$$\frac{dH}{dt} \approx \frac{mFnv^3}{2H_0a^3}. \tag{1}$$

Here $\underline{m}$ is the mass of atoms constituting the stream; F is the effective area of the magnetic "diameter" of the earth; $\underline{a}$ is the earth's radius; $\underline{n}$ is the concentration of atoms in a stream at the distance of the earth from the sun; H_0 is the horizontal component of the earth's magnetic field.

We can see from Equation (1) that the rate of change of the earth's magnetic field during the starting phase of a magnetic disturbance depends greatly on the velocity of the corpuscles in a stream. It is especially true, on the basis of the velocity data introduced above, that we can expect particularly fast variations in the earth's magnetic field to be observed at the intrusion of the corpuscles producing disturbances with a sudden commencement. From the same data on velocities and from Formula (1) it becomes obvious that among the weak disturbances, those of sudden commencement appear to be most intense, as was noted at the start of this article.

The significance of the atom concentration in the stream is the same as that of the velocity. Here however, into the various factors which are used in determining the magnitudes of $\underline{v}$ and $\underline{n}$ enters the product of these two magnitudes (or to be more exact, the product of $\underline{v}$ and $\underline{n}$ with various exponents). For this reason, up to this time we have not obtained any definite data pertaining to the differences in the magnitudes of $\underline{n}$ in the disturbances with a sudden and nonsudden commencement. Only the order of magnitudes $\underline{n}$ is known for the given velocities $\underline{v}$ (see [19]).

Let us pass to the third factor — the magnetic field of those condensations which compose the streams.

A great many factors indicate that such magnetic fields do exist in the streams moving away from the sun; the knowledge of this fact is specifically used in explaining some variations in the cosmic rays. Our basic problem here lies in determining which of the condensations have the stronger field — the ones which produce the disturbances with a sudden or those which produce the disturbances with a nonsudden commencement. It can be said, with a fair degree of certainty, that a substantially more intense field may be expected to exist in a case of those condensations which produce the disturbances with a sudden commencement. This can be deduced from the fact that these disturbances are related to the spots characterized by strong magnetic fields. Moreover, we have previously stated that there exists a basis for the belief that, in considering the emissions of the condensations producing such disturbances, we are dealing with a nonstationary process of electromagnetic nature. For this reason, we can expect that at the emission of the condensations from a given solar region, these condensations would carry with them noticeable magnetic fields "frozen" into the matter of the condensations. The part played by these "frozen" magnetic fields at the approach of the condensations to the earth should be considered as an auxiliary factor altering the earth's field at the time of a disturbance. Apparently, these magnetic fields may either intensify or weaken the variations of hydrogen, which are determined by intrusion of a condensation into the outer layers of the atmosphere. This problem calls for a special investigation.

---

*Reports of the Crimean Astrophysical Observatory; in press (Izvestiia Krymskoi astrofizicheskoi observatorii).

As to the condensations which produce disturbances with a nonstationary commencement, here the "frozen" magnetic fields must be less intense. This is proven by the fact that these disturbances are not related to the spots and, furthermore, they are observed at the time when sunspots are absent. Also, the disturbances with nonsudden commencements are not related to the intensifications of the solar radio emission, in which an important part is undoubtedly played by electromagnetic effects, etc. It follows from what has been said that the differences in a specific "frozen" field of the condensations may constitute a very important factor in the problem of explaining the differences in the character of the two types of the disturbances.

In conclusion we must note the following conditions. Even though division of the disturbances into those of sudden and those of nonsudden commencement appears, on the basis of the whole immense scope of material collected by now, to be necessary and well grounded, nevertheless such a division is difficult to carry out in many instances. Thus, for instance, in some cases the consecutive disturbances are characterized by sudden commencement, etc. These cases, however, represent exceptions to the general rule. Furthermore, these anomalous factors can be always explained if we take into consideration the fact that, in accordance with all that has been said in this article, the character of the disturbances depends on a great many conditions. For this reason we will refrain from conducting a special investigation of this problem.

Crimean Astrophysical Observatory,
Academy of Sciences, USSR

Received November 1, 1956

LITERATURE CITED

[1] H. Newton and A. Milsom, Journ. Geoph. Res. 59, 203, 214 (1954).

[2] E. Thellier and O. Thellier, Compt. Rend 227, 1044 (1948).

[3] L. Harang, The Aurora (1951), p. 15.

[4] C. Allen, Monthly Notices, Roy. Astron. Soc. 104, 13 (1944).

[5] H. Newton, Monthly Notices, Roy. Astron. Soc. Geophys. supplement 5, 321 (1948).

[6] E. R. Mustel', Izv. Krymskoi Astrofiz. Obs. 15, 95 (1955).

[7] E. R. Mustel', Doklady Akad. Nauk SSSR 42, 117 (1944).

[8] E. R. Mustel' Izv. Krymskoi Astrofiz. Obs. 3, 3 (1948).

[9] E. R. Mustel', Astron. J. (USSR 32, 177 (1955).

[10] C. Allen, Meterolog. Mag. 104, 13 (1944).

[11] H. Newton, Monthly Notices, Roy. Astron. Soc. 103, 244 (1943).

[12] S. K. Mitra, The Outer Atmosphere* (IIL, 1955).

[13] K. Kiepenheuer, Astrophys. J. 105, 408 (1947).

---

*Russian translation.

A POSSIBLE NEW TYPE OF FLUORESCENCE OF THE EARTH'S ATMOSPHERE

I. S. Shklovskii

As a result of cascade transitions during the resonance absorption of the solar emission line L_{β_3} by oxygen atoms of the earth's atmosphere (transition $^3P_2 - {}^3D$) infrared lines λλ 11299 - 11287 and λ 8446 will be emitted. This should lead to a twilight flash of these lines. The expected intensity of the infrared lines of oxygen during the twilight flash is calculated. For an undisturbed sun it is of the order of 10^8 photon/cm^2 sec. The line λ 8446 is in the region of the band (6, 2) of the vibration-rotation spectrum of hydroxyl in the earth's atmosphere. This makes the observations of the twilight flash of the oxygen line difficult but not impossible. The author emphasizes that systematic observations of this flash give a new possibility of investigating the emission lines of the Lyman series in the solar spectrum as a function of solar activity.

The distinct flash of the yellow sodium doublet in twilight is well known. We are also familar with the twilight flash of the red oxygen doublet λλ 6300-6364 A and of the λ 3914 A N_2^+ band; it is also possible that the oxygen band at λ 8600 A is also strengthened in twilight.

The lines and bands which have been mentioned are enhanced in twilight by resonance fluorescence. In the present article we shall discuss the possibility of twilight flash of certain infrared lines under the influence of ultraviolet radiation from the sun. This concerns the possibility of ordinary, rather than resonant, fluorescence of the earth's atmosphere.

It has been known for a long time that the ultraviolet spectrum of atomic oxygen contains a strong line whose wavelength is remarkably close to the wavelength of the familar hydrogen line L_β. For example, Hopfield [1] in a laboratory study of the OI spectrum pointed out the very intense line λ 1025.84 A which is close to L_β at 1025.73 A. The intensity of this oxygen line was estimated at 9 points on a ten-point scale. This line originates in the transition from the $^3D^0_{3,2,1}$ level to the ground level of oxygen, $^3P_2$. Later data [2] give the exact wave number of the $^3P_2$ - $^3D_0$ transition as $\nu_0 = 97\,488.14$ cm^{-1}, whereas for L_β we have $\nu_H = 97\,492.20$ cm^{-1}. The wave number difference between these two lines is only ~ 4 cm^{-1}, while the wavelength difference is $\Delta\lambda \approx 0.04$ A

Because of the extremely close wave numbers of the hydrogen and oxygen lines and the fact that the $^3P_2$ level of oxygen is lowest we can expect resonance effects in the earth's atmosphere due to solar emission in the L_β region. Let us consider this in greater detail. After resonance absorption of an L_β photon an atom of atmospheric oxygen rises to the $^3D^0_{3,2,1}$ state. It can then drop into the ground state $^3P_2$ or into $^3P_{2,1,0}$. In the first case, because of the large absorption coefficient at the center of the resonance line the emitted photon is immediately absorbed by an atom of atmospheric oxygen. This occurs in planetary nebulae. Therefore after L_β absorption the ground state $^3P_2$ of an oxygen atom will be reached preferentially by the cascade $^3D^0 \rightarrow {}^3P_{2,1,0} \rightarrow {}^3S \rightarrow$ $\rightarrow {}^3P$. In the $^3D^0_{3,2,1} \rightarrow {}^3P_{2,1,0}$ transition the infrared oxygen lines λλ 11294 and 11287 A will be emitted. The transition $^3P_{2,1,0} \rightarrow {}^3S_1$ gives the well known infrared line λ 8446 A which is characteristic of the auroral spectrum. The transition from $^3S_1$ to the ground level $^3P_2$ gives the ultraviolet line in which we are interested.

As a result of the fluorescence process which has been described every L_β photon that is absorbed by atmospheric oxygen yields the two infrared photons: $\lambda\lambda$ 11294-87 A and λ 8446 A. Thus we can expect from theory that the upper layers of the earth's atmosphere will fluoresce, emitting infrared oxygen lines, under solar irradiation.

We shall now discuss the intensity of this fluorescence and discuss the possibility of observing it. Rockets have been used to investigate solar ultraviolet radiation, especially L_α, at high altitudes. According to [3], the L_α flux from the undisturbed sun is 0.10 erg/cm$^2\cdot$sec and the line width is less than 1 A. The L_β line in the solar spectrum was recently observed [4] by rockets at 130-150 km, although no direct intensity data were given in this report. According to Goldberg's recent theory of the production of the L_α solar emission line [5] the intensity at the center of the line is determined by the Planck function for the kinetic temperature of the upper chromosphere layers where the optical density at the center is unity. This temperature as estimated from the L_α intensity [5] is 6000°. This is approximately the temperature which should exist in the region where L_β is emitted. The chromospheric layers where the Lyman lines originate is at about 4000 km. From this it can be calculated that the L_β flux should be about one order of magnitude smaller than for L_α. It is possible, however, that Goldberg's theory neglects some factors, so that the L_β flux may be larger. But for our further calculations we shall assume the L_β flux in solar radiation at the edge of the earth's atmosphere is $F_{L\beta} = 0.01$ erg/cm$^2\cdot$sec = $= 5\cdot 10^8$ photon/cm^2 sec, regarding this as a minimum.

We write the selective absorption coefficient of oxygen in the usual form

$$S_\nu = \frac{e^2}{mc}\,\frac{\gamma}{4\pi}\,\frac{f}{(\nu-\nu_0)^2+(\gamma/4\pi)^2}\,, \qquad (1)$$

where $\gamma = \frac{8\pi^2 e^2 \nu^2}{3\,mc^3} = 2.22\cdot 10^9\ \mathrm{sec}^{-1}$, from which, for $\nu-\nu_0 \gg \gamma/4\pi$,

$$S_\nu = \frac{1.53\cdot 10^6\cdot f}{(\nu-\nu_0)^2}\,. \qquad (2)$$

Considering the character of the $^3P_2 - {}^3D_{2,1}$ transition, the oscillator strength of the oxygen line $\lambda = 1025.83$ A must be quite large. According to the most recent data, in a column of unit cross section at atmospheric altitude $h_0 > 120$ km the number of oxygen atoms is given by $\int_{h_0}^{\infty} [\mathrm{OI}]\,dh = N_{h_0} = 7\cdot 10^{17}\ \mathrm{cm}^{-2}$. At $h_0 = 130$, 150 and 175 km, respectively, $N_{h_0} = 3\cdot 10^{17}\ \mathrm{cm}^{-2}$, $10^{17}\ \mathrm{cm}^{-2}$ and $2.5\cdot 10^{16}\ \mathrm{cm}^{-2}$. When $\nu - \nu_0 = 1.42\cdot 10^{11}\ \mathrm{sec}^{-1}$ (which corresponds to $\Delta\lambda = 0.05$ A) according to Equation (2), $S_\nu/f = 0.76\cdot 10^{-16}\ \mathrm{cm}^2$; for $\Delta\lambda = 0.10$ A, $S_\nu/f = 1.89\cdot 10^{-17}\ \mathrm{cm}^2$, and for $\Delta\lambda = 0.15$ A, $S_\nu/f = 8.4\cdot 10^{-18}\ \mathrm{cm}^2$. We can therefore conclude that at 175 km the absorption of solar L_β by atmospheric oxygen cuts out a band of width ~ 0.1 A (if f ~ 1) in the L_β line; at 150 km the width is ~ 0.2 A; at 130 km it is ~ 0.3 A and at 120 km it is ~ 0.5 A. The width of L_β must be less than 0.8 A, since the width of L_α is less than 1 A Therefore at $h_0 = 175$ km we have absorption (or more precisely, scattering) of more than 1/8 of the flux of solar L_β, at $h_0 = 150$ km more than 1/4, and at 120 km more than 1/2. When f ~ 0.01 the width of this band is ten times less and the relative number of scattered photons is diminished in the same ratio. Hereinafter we shall assume f ~ 1.

Thus in a column of unit cross section in twilight every second there will be produced from $7\cdot 10^7$ to $3\cdot 10^8$ infrared photons at $\lambda\lambda$ 8446 and 11294-87 A for $175 > h_0 > 120$ km. The intensity of the oxygen line λ 6363 A in twilight flash is ~ $3\cdot 10$ photon/cm^2 sec. We can thus conclude that the theoretical intensity of infrared oxygen lines in twilight flash is not large but still sufficient to raise the question of their observation. Such observations, which can be very important (see below), encounter several great difficulties. We shall, of course, first consider the shorter wavelength λ 8446 A, because modern radiation receivers (such as electron-optical image converters) have much less sensitivity at λ 11300 A than at λ 8500 A. In the region of the oxygen line λ 8446 A we find the intensity (6, 2) rotation-vibration band of atmospheric hydroxyl. The emission of the night sky in this spectrum region can be described by Table 1 [6].

TABLE 1

λ Å	I (6308 = 8)	Identification
8418.0	11	(6,2) P_2
8432.3	44	(6,2) P_1
8446.0	(5—20)	0I
8455.2	11	(6,2) P_3
8467.4	35	(6,2) P_3
8497.3	7	(6,2) P_4

The characteristics of the suggested oxygen line λ 8446 A in the twilight flash spectrum are shown in parentheses. Its intensity is based on the assumption that the intensity of λ 6300 A in the night sky spectrum is 10^8 photon per $cm^2 \cdot sec$. Intensities 5-20 are assumed to cover the anticipated range for different solar angles. We can thus expect that in twilight flash the intensity of the oxygen line will be about the same as in weak satellites of P branch lines in the (6, 2) hydroxyl band. But we must keep in mind that in twilight the exposure time is limited and the brightness of the sky is quite high. All of these circumstances interfere seriously with observation of the twilight flash of λ 8446 A but do not make it impossible. The sought line is 14 A from the strong P_2 component and 11 A from the weak P_2 component of the (6, 2) band. Therefore successful observation would be assured by a spectrograph of sufficiently high dispersion and resolving power. The sensitivity of the receiver must also be very high. Such spectrographs are known to exist at the present time.

The foregoing calculations for the intensity of infrared oxygen lines in twilight flash refer to an undisturbed sun. During large flares and perhaps other manifestations of solar activity the L_β flux may be increased by a factor of a few tens. On the other hand, the fluctuations of the L_β and L_α solar lines must be synchronous. Therefore, the intensity of the oxygen lines of the earth's atmosphere in twilight flash would be a very sensitive, practically ideal indicator of the solar ultraviolet flux in Lyman lines.

There are strong reasons for believing that Lyman radiation is of great geophysical importance. For example, it is possible that one cause of the formation of the D layer is ionization of NO molecules by L_α. Therefore during the course of the International Geophysical Year it is proposed to make a special systematic search by sending up rockets with spectrographs during flares. If it becomes possible to detect twilight flash of oxygen lines we shall have a very efficient as well as a relatively inexpensive and simple method of studying the Lyman lines in solar radiation.

Institute of Atmospheric Physics,
Academy of Sciences, USSR

Received July 20, 1956

LITERATURE CITED

[1] J. Hopfield, Astrophys. J. 59, 114 (1924).

[2] Ch. Moore, Atomic Energy Levels (Washington, 1949).

[3] E. Byram, T. Chubb, H. Friedman and N. Gailar, Phys. Rev. 91, 1278 (1953).

[4] F Johnson, H. Malitson, J Purcell and R. Tousey, Astron. J. 60, 165 (1955).

[5] L. Goldberg, Astrophys. J. 120, 185 (1954).

[6] A. Meinel, Astrophys. J. 111, 555 (1950).

OBSERVATIONS OF HYDROGEN EMISSION IN AURORAE

Iu. I. Gal'perin

Data are given on 17 spectrograms containing H_α, taken during observations at the magnetic zenith and the magnetic horizon. Profiles of H_α have been determined from the three best spectrograms taken at the zenith. These give a maximum value of the velocity of approach of the order of 2000 km/sec. The velocity corresponding to maximum density, equals 350 ± 100 km.

Investigations over many years by Vegard [1] and Gartlein [2], who used low-dispersion spectrographs, have proved that aurorae contain broadened Balmer lines of hydrogen. Spectrograms taken by Meinel [3] and Gartlein [4] at the magnetic zenith and horizon, with high dispersion and high resolving power, have provided convincing evidence that protons move approximately along the magnetic lines of force of the earth's field with velocities of the order of thousands of kilometers per second. This has also been confirmed by Vegard [1].

However, to our knowledge, besides these spectrograms which took the magnetic field into account, the literature contains no other similar hydrogen line profiles with sufficient resolution.

During the period from December 1955 to April 1956 at the northern station of the Institute of Atmospheric Physics of the Academy of Sciences USSR, which is located in the auroral zone ($\varphi = 68° 38'$; $\lambda = 2^h 13^m.3$), observations were made for the purpose of preparing to obtain profiles of hydrogen lines in the aurora during the International Geophysical Year.

This work was done with an SP-48 GOI spectrograph,* of 1 : 0.8 aperture ratio and 83.5 A/mm dispersion for H_α and resolution limit about 2 A, using pre-exposed Panchrome X film [6]. During this period 17 spectra were obtained containing $H\alpha$, of which two were taken at the magnetic horizon and the others at the magnetic zenith (more precisely, 8° north of the magnetic zenith). The magnetic horizon photographs were taken without reinforcing optics (the collimator aperture ratio was 1 : 4.7). In the magnetic zenith direction an arc of 4° was projected on the spectrograph slit.

The H_α line was obtained from a quiet HA arc, from diffuse DS and PS emission and from coronal forms C and D (this notation for the forms is taken from [7]).

All of the hydrogen lines belonged to weak and average aurorae. Table 1 gives the K index of magnetic disturbances during exposure averaged for all observatories of the Soviet Union; the K index is seen not to have exceeded 5.

The H_α intensity varies strongly compared with the intensity of the first positive N_2 system. In some spectrograms (Nos. 1, 2, 14, 15) H_α is pronounced but the first positive nitrogen system is almost absent. There are some instances of the reverse. An absolute calibration was not performed, but it can be stated that the H_α intensity did not once exceed the intensity of [OI] λ 6363.8 A.

---

* This spectrograph is of the type described in [5].

TABLE 1

No.	Date	Exposure (Greenwich time)	Auroral form	K index	Geomagnetic coordinates on celestial sphere
	Dec. 1955				
1	17	16^h30^m—20^h30^m	HA	2	Magnetic horizon
2	17—18	21 30 —01 30	DS. HA	2	
	Feb. 1956				
3	5	20 00 —22 00	PA. C. DS, PS	4	Magnetic zenith
4	5	20 00 —23 10	HB. HA. PS	3	
5	11	16 45 —19 10	HB. HA. RA	5	
6	11	19 10 —19 30	HA. RA. C	5	
7	11—12	22 20 —02 30	RA. C. PS	5	
8	12—13	17 15 —00 15	HA. PS	3	
9	13	00 15 —02 00	HA. C. PS	2	
10	16	16 45 —18 15	HA	4	
11	16	18 15 —20 30	HB. RB. PS	3	
12	27	20 00 —23 30	HA	5	
	Mar. 1956				
13	6	17 30 —19 30	HB. C. DS	4	
14	11	19 15 —21 40	HA	3	
15	11	21 40 —22 40	CD	3	
16	11—12	22 40 —02 30	C. PS	3	
17	16—17	21 00 —00 40	HB. RB	2	

TABLE 2

No.	Velocity in km/sec			
	v_A	v_B	v_C	v_D
14	+ 250	− 300	− 1700	− 1850
15	+ 350	− 350	− 1700	− 1950
16	+ 300	− 400	− 1950	− 2050

A comparison with solar data [8] shows that in all cases about one day before exposure a calcium flocculus was located within the interval 8-15° from the point of projection of the earth on the solar disk.

Some spectrograms, particularly Nos. 1 and 2, were distorted by nonuniform pre-exposure. The most reliable H_α profiles at the magnetic zenith were obtained from three well developed spectra (Nos. 14, 15, 16) with the highest ratio of H_α intensity to the intensity of the interfering bands of the first positive N_2 system.

Figure 1 shows traces of spectra Nos. 14, 15, 16 obtained with an MF-4 microphotometer (21 × magnification, microphotometer slit width 0.65 mm, resolution limit $\underline{r}$ 3 A). For comparison Figure 2 shows Meinel's traces (resolution limit 7 A) and Gartlein's traces (resolution limit about 15 A).

In order to obtain profiles the trace of the faint background of the first positive N_2 system was subtracted from the H_α trace. Three different background traces were subtracted and the results were averaged. Figure 3 shows the profiles Nos. 14, 15 and 16 obtained in this manner. Densities were not converted to intensities. These profiles permit us to determine the velocities of protons (v_A, v_B, v_C, v_D) corresponding to the following points of the profile: A is the maximum observed velocity of recession from the observer ("red shift"); B is the maximum density; C is the density 0.1 of maximum; D the maximum observed velocity of approach to the observer ("violet shift"). These velocities are given in Table 2.

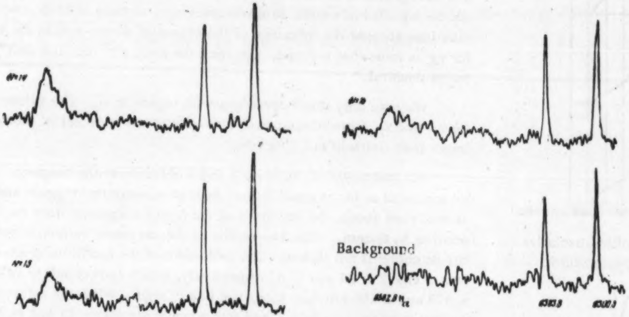

Fig. 1. Traces Nos. 14, 15 and 16 and background without H_α.

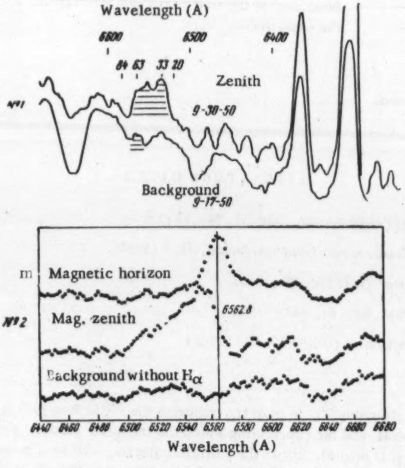

Fig. 2. Traces obtained by Gartlein (No. 1) and Meinel (No. 2).

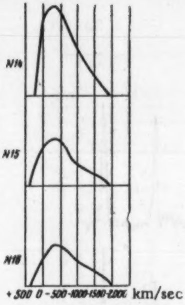

Fig. 3. H_α profiles in relative densities. Magnetic zenith.

The error in determining velocity v_B is ± 100 km/sec; for v_A, v_C and v_D the error is ± 200 km/sec. The result obtained for the maximum observed proton velocity of approach v_D depends on the film sensitivity and background fluctuations; therefore the values obtained for v_D must not be regarded as exact. In this respect v_C was more reliable. When we take into account the influence of the apparatus on the profile the value for v_A is somewhat reduced; therefore the reality of the "red shift" remains doubtful.*

We note only small deviations with regard to v_B. The values obtained for v_C and v_D were smaller than those obtained by Meinel (K index 9) but larger than Gartlein's (K index 5).

For spectrograms numbers 1 and 2 obtained at the magnetic horizon we excluded as far as possible the effect of nonuniform exposure and also, as described above, the influence of the faint background from the first positive N_2 system. The H_α profile at the magnetic horizon is broadened but its center is not shifted. The half width of the profile in spectrograms Nos. 1 and 2 is 14 and 17 A, respectively, which corresponds to velocities ± 300 and ± 400 km/sec along the line of sight. When the effect of the apparatus on the profile is taken into account we obtain 11 and 14 A, corresponding to ± 250 and ± 300 km/sec. As a result of nonuniform exposure the accuracy of the profile obtained from these spectrograms is less than for Nos. 14, 15 and 16.

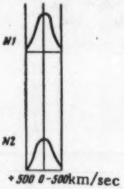

Fig. 4. H_α profiles in relative densities. Magnetic horizon.

The author is deeply grateful to V. I. Krasovskii for his continued interest and direction of this work and to M. A. Ermolaev and V. F. Musatova of the staff of the northern station for their assistance with the experimental work.

Institute of Atmospheric Physics,
Academy of Sciences, USSR

Received August 11, 1956

LITERATURE CITED

[1] L. Vegard, Geofys. Publikasjoner, Oslo. 18, No. 15 (1952).

[2] C. W. Gartlein, Trans. Amer. Geophys. Union, 31, 7 (1950).

[3] A. B. Meinel, Astrophys. J. 113, 50 (1951).

[4] C. W. Gartlein, Phys. Rev. 81, 463 (1951).

[5] Iu. I. Gal'perin, Astron. J. (USSR) 33, 173 (1956).

---

*As was first noted by I. S. Shlovskii [9], in high H_α emission the mean free path is much smaller than the radius of curvature of the spiral, and the time during which the incoming particle is charged is much shorter than the time during which it is neutral, so that the particle trajectory will be a broken line rather than a spiral. It appears to us that the "red shift" may be simply the result of particle deflection during multiple elastic scattering.

[6] Sensitometry Handbook (GITTL, Moscow, 1955) [in Russian].

[7] Photographic Atlas of Auroral Forms (Oslo, 1951).

[8] Solar Data [in Russian, Biulleten' Akad. Nauk SSSR No. 6 (1955); Nos. 1-2 (1956).

[9] I. S. Shklovskii, Doklady Akad. Nauk SSSR 81, 367 (1951).

BRIEF COMMUNICATIONS

ON THE UTILIZATION OF A THIN LUMMER-GEHRKE PLATE

L. A. Fedin

A thin Lummer – Gehrke plate is described. The formula for the resolving power is derived and calculated parameters of the plate are given. These show that the conclusions reached by A. V. Merkulov in his paper "A Stellar Spectrograph with a Lummer – Gehrke Plate" are erroneous.

Characteristics of a thin Lummer – Gehrke plate and considerations showing the inadvisability of its utilization are given.

A. V. Merkulov [1] has discussed a thin Lummer-Gehrke plate, of the order of 0.2 mm, placed in optical contact with a thick glass base. Table 1 shows data obtained by Merkulov and by the present author for the H_α line.

TABLE 1

	Quantity	Merkulov's data	Data of present author
Initial data	Material of plate	Flint L-23	Flint TF4 $n_C = 1.73243$
	Material of base plate	Crown L-20	Crown K3 $n_C = 1.50760$
	Thickness t of plate	0.2 mm	0.2 mm
	Length L of plate	25.4 mm	30 mm
	Angular dispersion $\frac{d\epsilon}{d\lambda}$	0.004 rad/A	0.016 rad/A
Calculations	Spectrum order m	492	520
	"Plate constant" g	12.10 A	11.48 A
	Number of interfering beams N	34	43
	Resolving power R	$6.1 \cdot 10^7$	$2.43 \cdot 10^4$

Small differences in the results can be attributed to the different kinds of glass employed. However, the difference in the resolving power by a factor of more than a thousand results from Merkulov's error in deriving the formula for the resolving power (see [1], p. 151, Formula (16)).

A correct derivation can be obtained by appropriately modifying the derivations given by Frish [2] or Tolanskii [3].

Let n_1 be the refractive index of the plate and n_2 the refractive index of the base, with $n_1 > n_2$. Denoting the thickness of the plate by t (Fig. 1), we obtain for the difference between two successive beams

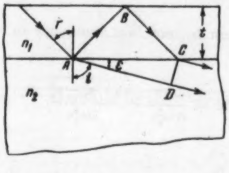

Fig. 1.

$$\Delta = n_1(AB + BC) - n_2 AD = \frac{2tn_1}{\cos r} - 2tn_2 \tan r \sin i.$$

Using $n_2 \sin i = n_1 \sin r$, we obtain

$$\Delta = 2tn_1\left(\frac{1}{\cos r} - \tan r \sin r\right) = 2tn_1 \cos r.$$

By substituting $\cos r = \sqrt{1 - \sin^2 r} = \sqrt{1 - \frac{n_2^2}{n_1^2}\sin^2 i}$, we obtain

$$\Delta = 2t\sqrt{n_1^2 - n_2^2 \sin^2 i}.$$

The condition for the formation of a principal maximum is

$$m\lambda = 2t\sqrt{n_1^2 - n_2^2 \sin^2 i},$$

or, replacing the angle i by the Bragg angle $\epsilon = \frac{\pi}{2} - i$:

$$m\lambda = 2t\sqrt{n_1^2 - n_2^2 \cos^2 \varepsilon}. \quad (1)$$

This expression is satisfied by the angle ϵ at which an m-th order principal maximum is formed; the angle $\epsilon + \Delta\epsilon$ corresponding to the (m + 1)-th order for the same wavelength satisfies the condition

$$(m + 1)\lambda = 2t\sqrt{n_1^2 - n_2^2 \cos^2(\varepsilon + \Delta\varepsilon)}. \quad (1a)$$

The order is determined from Equation (1):

$$m = \frac{2t}{\lambda}\sqrt{n_1^2 - n_2^2}. \tag{2}$$

Squaring (1) and (1a), subtracting term by term, and assuming ϵ to be small, we obtain the angular separation of two successive orders:

$$\Delta\varepsilon = \frac{m\lambda^2}{4t^2 n_2^2 \varepsilon} = \frac{\lambda\sqrt{n_1^2 - n_2^2}}{2tn_2^2\varepsilon}. \tag{3}$$

Differentiating (1), we obtain the angular dispersion:

$$\frac{d\varepsilon}{d\lambda} = \frac{n_1^2 - n_2^2 - \lambda\left(n_1\frac{dn_1}{d\lambda} - n_2\frac{dn_2}{d\lambda}\right)}{\lambda n_2^2\varepsilon}. \tag{4}$$

Hence

$$d\lambda = \frac{\lambda n_2^2\varepsilon}{n_1^2 - n_2^2 - \lambda\left(n_1\frac{dn_1}{d\lambda} - n_2\frac{dn_2}{d\lambda}\right)}\,d\varepsilon. \tag{5}$$

Replacing $d\epsilon$ by the angular separation $\Delta\epsilon$ of two successive orders according to (3), we obtain the separation of successive maxima expressed in wavelengths (the so-called "plate constant"):

$$g = \Delta\lambda = \frac{\lambda^2\sqrt{n_1^2 - n_2^2}}{2t\left[n_1^2 - n_2^2 - \lambda\left(n_1\frac{dn_1}{d\lambda} - n_2\frac{dn_2}{d\lambda}\right)\right]}. \tag{6}$$

The wavelength difference $\delta\lambda$ between two lines that can barely be separated with N interfering beams is

$$\delta\lambda = \frac{\Delta\lambda}{N} = \frac{\lambda^2\sqrt{n_1^2 - n_2^2}}{2tN\left[n_1^2 - n_2^2 - \lambda\left(n_1\frac{dn_1}{d\lambda} - n_2\frac{dn_2}{d\lambda}\right)\right]}. \tag{7}$$

For plate length L the number of interfering beams is

$$N = \frac{L}{AC} = \frac{L}{2t\cdot\tan r}.$$

Since the angle r is very close to the angle of total internal reflection, we can assume approximately

$$\frac{n_1}{n_2}\sin \;= 1 \text{ and } \tan r = \frac{1}{\sqrt{\frac{n_1^2}{n_2^2} - 1}},$$

hence

$$N = \frac{L\sqrt{\frac{n_1^2}{n_2^2} - 1}}{2t}. \qquad (8)$$

Using Formulas (7) and (8), we obtain an expression for the resolving power:

$$R = \frac{\lambda}{\delta\lambda} = \frac{L\sqrt{\frac{n_1^2}{n_2^2} - 1}\left[n_1^2 - n_2^2 - \lambda\left(n_1\frac{dn_1}{d\lambda} - n_2\frac{dn_2}{d\lambda}\right)\right]}{\lambda\sqrt{n_1^2 - n_2^2}}. \qquad (9)$$

The solution of Hartmann's dispersion formula $n = n_0 + \frac{c}{\lambda - \lambda_0}$ which is required for our calculation, gives for K3 glass the following constants: $c = 76.6615$, $n_0 = 1.49218$, $\lambda_0 = 1595.74$ A, and differentiation of the formula gives the dispersion for the red hydrogen line $\lambda_C = 6562.8$ A:

$$\frac{dn_2}{d\lambda} = -3.1073 \cdot 10^{-6}.$$

For TF4 glass we obtain the following values: $c = 162.317$, $n_0 = 1.69382$, $\lambda_0 = 2359.86$ A, $\frac{dn_1}{d\lambda} = -9.1918 \cdot 10^{-6}$. Thus such a thin Lummer-Gehrke plate loses its principal advantage, that of high resolving power, because prism spectrographs possess resolution of the order of $2 \cdot 10^4$.

The dispersion of this plate is almost indistinguishable from the dispersion of a prism spectrograph, with which Merkulov compared the plate. The plate retains one advantage over a prism spectrograph, its higher transmission.

The plate has another shortcoming by comparison with a prism spectrograph: the interval of the spectrum which can be studied does not exceed 10-15 A.

Moreover, it is very difficult to prepare such a thin plate with excellent surfaces (to 1/80 of a wavelength, as shown by Tolanskii [4]) in optical contact with a base plate. It must also be kept in mind that a plate is very sensitive to temperature fluctuations and requires thermostatic control to within a few hundredths of a degree (Korolev [5]).

It is thus clear that Merkulov was in error in asserting the advantage of a thin plate, and the advisability of using it is at least doubtful.

Received October 15, 1955

LITERATURE CITED

[1] A. V. Merkulov, Astron. J. (USSR) 23, 149 (1946).

[2] S. E. Frish, Spectroscopic Techniques [in Russian] (Leningrad, 1936) p. 82.

[3] S. Tolanskii, High Resolution Spectroscopy [in Russian] (1953) p. 212.

[4] S. Tolanskii, High Resolution Spectroscopy [in Russian] (1953) p. 241.

[5] F. A. Korolev, High Resolution Spectroscopy [in Russian] (1953) p. 88.

A NUMERICAL METHOD FOR DETERMINING THE STELLAR DENSITY IN A SPHEROIDAL CLUSTER

A. S. Sharov

P. N. Kholopov recently worked out and successfully applied a numerical method for determining the spatial density of stars in a spheroidal cluster [1]. In this method it is assumed that the cluster consists of n spheroidal shells of equal density, formed by concentric, similar, and similarly arranged oblate spheroids, with the ratio of the major to the minor semi-axis represented by η. The principle of this method, first developed by Wallenquist [2] for spheres, is as follows. We calculate the volume of that part of the outermost spheroid which in projection on the celestial sphere gives the first outer elliptical ring. By counting the number of stars in the ring, we find the mean density in the layer. Then we calculate the volumes of those parts of the second and first layers which give in projection the second more inward ring. By counting the stars in the second ring and knowing their density in the first shell and also the volume of the region in which they are found, we obtain the density in the second shell. Proceeding to the center, we can determine the stellar densities in all layers. The essence of this method is the calculation of the volumes of the corresponding parts of the spheroidal shells. For the calculation of these volumes it is possible to suggest a method based on the theory of affine transformations. This method is much simpler than the one applied by Kholopov.

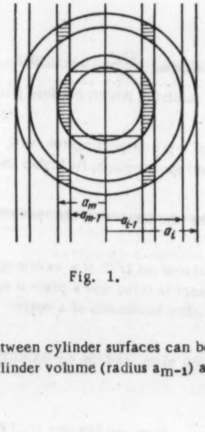

Fig. 1.

Let us consider first a spherical cluster in which spheres with radii a_{m-1}, a_m and a_{i-1}, a_i form any two shells. We describe around the spheres with radii a_{m-1} and a_m two right cylinders with generating lines parallel to the line of sight. In Figure 1 a cross section of such a cluster is given. That part of the volume of the shell (a_{m-1}, a_m) lying between cylinder surfaces can be calculated as the difference between the sphere volume (radius a_m), and the cylinder volume (radius a_{m-1}) and two polar caps. Well known formulas of elementary geometry give

$$V = \frac{4}{3}\pi a_m^3 - 2\pi a_{m-1}^2 \sqrt{a_m^2 - a_{m-1}^2} -$$

$$- \frac{2}{3}\pi \left(a_m - \sqrt{a_m^2 - a_{m-1}^2}\right)^2 \left(2a_m + \sqrt{a_m^2 - a_{m-1}^2}\right)^2 = \frac{4}{3}\pi (a_m^2 - a_{m-1}^2)^{3/2}.$$

The volume formed by the same cylinders in the shell (a_{i-1}, a_i) is obviously:

$$W = \frac{4}{3}\pi \{(a_i^2 - a_{m-1}^2)^{3/2} - (a_i^2 - a_m^2)^{3/2} - [(a_{i-1}^2 - a_{m-1}^2)^{3/2} - (a_{i-1}^2 - a_m^2)^{3/2}]\}.$$

We now consider a spheroidal cluster and apply the theory of affine transformations. As we know, a spheroid is an affine transformation of a sphere and can be obtained from it by contraction along the coordinate axes and a rotation. Thus all volumes change proportionally to the product of the contraction coefficients. Let the spherical cluster contract along the line of sight with a coefficient η. Then the volumes V and W change by a factor η. Thus, the formulas for the volumes of interest in the case of a cluster contracted along the line of sight differ from the preceding formulas only by the factor η. Here a_1 is the major semi-axis of the contracted cluster. The formulas do not change further if the minor semi-axis of the cluster makes some angle $\underline{i}$ with the line of sight. In fact, all volumes change in an affine contraction by a coefficient η, independently of the angle formed by the cylinder axis with the line of sight. We select this angle in advance in such a way that by contraction of the cluster along the ultimate minor axis (which makes an angle $\underline{i}$ with the line of sight) the cylinder axis will fall along the line of sight. In this way, the formulas for the considered volumes are independent of the angle $\underline{i}$ and the geometric picture created by this transformation will correspond to the real conditions of observation of a spheroidal cluster.

P. K. Shternberg State Astronomical Institute

Received October 15, 1956

LITERATURE CITED

[1] P. N. Kholopov, Astron. J. (USSR) 26, 298 (1949).

[2] A. Wallenquist, Bosscha Ann. 4, 5 (1933).

CHRONICLE

SEVENTH CONFERENCE ON METEORITES

From the 14th to the 17th of November, the Seventh Conference on Meteorites was held in the House of Scientists of the Academy of Sciences of the USSR, convened by the Committee on Meteorites of the Academy. The presidency of the Academy of Sciences had invited to the conference foreign scientists, who took part in its work.

The Conference was opened with a report by V. G. Fesenkov on "Progress of the study on meteorites." The author cited the most outstanding works in the field of meteorites during the interval after the sixth Conference, and described the great role of this study in the solution of a number of important cosmogenic problems.

The report by A. A. Iavnel' on "Regularities in the composition of meteoritic material and the conditions of formation of meteorites" aroused great interest among the members and gave rise to a lively discussion. Thorough analysis of data principally gathered from chemical analysis of meteoric material, including mineralogical and petrographic characteristics of meteorites, allowed the author to find features common to all varieties of meteoritic material. The division of meteorites according to these features into 5 groups and 6 sub-classes, sufficiently different from each other, made it possible for the author to draw a number of far-reaching conclusions, particularly concerning the origin of the whole complex of meteoritic material coming from five asteroids having a layered structure. The regular variations in the chemical composition of meteorites, and consequently of the asteroidic bodies, leads to the concept of a common origin of these asteroids (and possibly comets) in an original proto-planetic cloud.

V. G. Fesenkov touched, in his report on "Some general problems of meteorics and astronomy," upon a number of problems connected with the origin of lunar craters, asteroids, and meteorites in the light of contemporary concepts of the origin of the solar system. Many data of investigation of lunar forms are more in favor of the volcanic than of the meteoric hypothesis of their origin. In any case, there remains the given problem, which is still open to discussion. Investigations of meteorites from the point of view of their luminosity, color, period of revolution, and forms make it possible to conclude that the asteroids are fragments of greater bodies and cannot have originated from the destruction of only one parent planet, which is confirmed by the different periods of revolution of the asteroids and by the division of meteorites into certain groups according to their chemical composition (Urey and Craig, Iavnel'). The origin of asteroids and consequently of meteorites is connected with the conceptions of the initial material from which our solar system originated. In this connection, the problem of the origin of the chondrulic structure of stony meteorites is a basic one, in the consideration of which the author came to the conclusion that the most probable hypothesis for the origin of the chondrulic structure is "that a similar structure was already inherent in the planet itself and was inherited by it from the original gas-dust nebula formed simultaneously with the material of the planet-asteroid itself."

Later on, Prof. N. Bonev (Bulgarian National Republic) in his lecture on "A meteoric hypothesis of the origin of lunar craters," starting from definite premises on the formation of craters as a result of the fall of meteorites, proved mathematically by a statistical calculation of the distribution of lunar forms between its eastern and western hemispheres that the craters are not related to a meteoric origin, at any rate not substantially.

The whole morning session of the second day of the Conference was devoted to papers on the Sikhote-Alin* iron meteorite. A paper was read by N. B. Divar' on "Final elements of the atmospheric trajectories of the Sikhote-Alin' meteoritic rain," in which more precise data were given on the trajectories of the meteorite in the terrestrial atmosphere. In his report, "The pattern of the fall of the Sikhote-Alin' meteoritic rain," E. L. Krinov gave an exhaustive picture of the fall of the meteorite on earth, obtained by an analysis of the distribution of the meteoritic masses and the character of the craters formed. From a detailed review of the characteristics of the ellipse of scattering of the meteoritic rain, a picture was established of the motion and disintegration of the meteoritic body in the atmosphere.

---

* Transliteration of Russian — Publisher's note.

The paper on "Chemical composition of the Sikhote-Alin' meteorite" read by M. I. D'iakonova was devoted to the results of chemical analysis of the meteorite. Because of an apparent homogeneity in the composition, it was possible to determine the total composition of all the meteoritic material and also the average chemical composition of nickel-iron of some samples of the meteorite. The total nickel content in the meteorite was shown to be 5.94% and in nickel-iron 6.00% (by weight). In the second part of the paper, data were given of chemical analyses of some of the most abundant minerals in the meteorite: troilite FeS, schreibersite $(Fe, Ni)_3P$, chromite $(Fe, Mg)Cr_2O_4$, kamacite (α-iron), together with a short description of the methods used for their determination.

L. G. Kvashi in his report on "Mineral composition and structure of the Sikhote-Alin' iron meteorite" gave detailed information on the petrographic characteristics and the mineral composition of the meteorite.

Starting from the chemical composition of separate minerals, based on their quantitative mineralogical analysis, the total composition of the meteorite was calculated and was shown to be approximately the same as the average composition determined by chemical analysis. The mineral composition of the meteorite consisted principally of kamacite (α-iron), taenite (γ-iron) and the additional minerals troilite, schreibersite, chromite, etc. From a detailed review of the structural characteristics, it was established that the Sikhote-Alin' meteorite belongs to a group of very coarse-structured octahedrites. From the data of petrographic and mineralogical investigations of the melting crust and of meteoritic dust of the Sikhote-Alin meteorite, the authors V. D. Kolomenskii and I. A. Iudin came to the conclusion that the crust has a double-layered structure and its outer layer mainly consists of magnetite. The meteoritic dust also consists of magnetite.

Here the series of problems concerning this fall ended.

The participants of the Conference watched with interest the scientific documentary film "The Sikhote-Alin' meteorite."

The report of E. K. Gerling and L. K. Levskii on "The origin of inert gases in stony meteorites" aroused general interest. The isolation from some meteorites of the inert gases Ar, Ne and He and the determination of their isotopic composition led to the following conclusions: the large Ar^{40} content in meteorites is connected with the radioactive decay of K^{40}; the ratio Ar^{36}/Ar^{38} which was established and the isotopic composition of Ne in meteorites differ from those in air; the recrystallized chondrules contain the same isotopes of the inert gases as the original ones, which is found to be a result of the accumulation of these isotopes after the recrystallization of the meteoritic material caused by nuclear reactions with cosmic rays. The large quantity of inert gases, as compared with other meteorites, and the different isotopic composition of the meteorite "Staroe Pes'ianoe" indicate that the latter meteorite once belonged to a very large cosmic body.

In the report of M. M. Shats on the work of I. E. Starik and M. M. Shats, who, by two parallel but different methods, deterined the uranium content of meteorites and isotopic composition, he drew the conclusion that the uranium content in meteorites is generally lower than on earth, while the isotopic composition is the same. The data obtained by them showed again the constant occurrence of uranium in meteorites.

In the lecture of K. P. Staniukovich and N. I. Shchetinina "On the problem of the inpact of a solid at high velocity on a compact medium" the character of shockwave propagation in a compact medium was considered, as well as the determination of the velocity, density, and pressure dependent on the impact-center distance.

Some considerations based on experimental material concerning the formations of chondrules from the gas phase, in a way similar to colloidal particles, were put forth in the paper of B. Iu. Levin and G. L. Slonimskii "On the problem of the origin of chondrules."

In a report by A. Aaloe, the results of geological research and excavation of craters on the island of Saarema (Estonia) were given and these data fully confirmed the meteoritic nature of the craters. About 0.5 kg of meteoritic matter was collected. The study of the structural characteristics and the composition of meteoritic dust particles found in the region of the crater were of particular interest for the investigators.

The paper by E. S. Burkser and K. N. Alekseeva was devoted to the physical properties of meteorites, to which little study has yet been given. The results of the study of some physical properties of stony meteorites showed that, as compared to the peridotites and pyroxenites which have about the same composition as stony meteorites, the latter have a lower heat capacity than terrestrial crystallized rocks, a higher specific gravity and a higher electroconductivity, which can be explained by its nickel-iron content. On this content depend also the magnetic properties of meteorites. Stony meteorites have a relatively low melting point and high porosity, caused by their particular chondrulic structure. On the basis of the results of investigation of the above-mentioned physical properties, an attempt was made to draw a picture of the origin of meteorites from asteroids.

I. A. Iudin reported the results of petrographic investigations of some nontransparent minerals of primary and secondary origin which he had observed in stony meteorites, e.g., nickel-iron, non-nickel iron, native copper, troilite, iosite, ilmenite, chromite. Their structural features were studied in detail.

In his paper concerning the petrographic investigations of the "Nikol'sk" stony meteorite which fell in 1954, L. G. Kvashi characterized this meteorite as a typical chondrite, distinguished by the abundance of chondrules of diameters varying from microscopical to 5 mm, and also by the special porosity of the basic mass, which contained about 10% metallic iron.

The x-ray analysis, considered by V. I. Mikheev, is based on a definite dependence of the variation of crystal-lattice parameters on the mineral composition, which gives it a wide application for minute analysis of the composition of meteorites.

F. Heide (German Democratic Republic) considered the origin of the craters of Carolina Bays. From aerial photographs and observations, F. Heide believed these craters to have been originated by the action of the sea on dry land. There are similar forms in the USSR near the town of Osipenko.

G. S. Pokshivnitskii (Poland) reported on meteorites in Poland and on new findings of the iron meteorite "Morasko."

The Seventh Conference on Meteorites concluded its work with a discussion of the papers and information given from the floor.

The very wide range of problems connected with the appearance and falling of meteorites and with the study of their composition, clarified at the Conference, illustrates the great interest among the scientists both in meteorites and in cosmic material, the investigation of which will contribute to the solution of a series of important geochemical and cosmogonic problems.

A display of literature on meteorites was arranged for the participants in the Conference.

M. D'iakonova

Received December 7, 1956

PUBLISHED BY THE ACADEMY OF SCIENCES OF THE USSR

The following books are available:

Astronomical Yearbook of the USSR, 1957, vol 36 (Institute of Theoretical Astronomy) 1954. 584 pp. Price, bound: 67 roubles, 50 kopecks.

Astronomical Yearbook of the USSR, 1958, vol. 37 (Institute of Theoretical Astronomy) 1955. 584 pp. Price, bound: 67 roubles, 70 kopecks.

Bredikhin, F. A. Studies on Meteorites, The Essays and Comments of A. D. Dubiago. Editor, S. V. Orlova (Series "Classics of Science"). 1954, 607 pp. 5 Appendixes. Price, bound: 24 roubles, 20 kopecks.

Problems on cosmogony Part. II. 1954. 364 pp. Price, bound: 17 roubles, 35 kopecks.

Problems on cosmogony Part. III. 1954. 363 pp. Price, bound: 16 roubles, 60 kopecks.

Problems on cosmogony Part IV. 1955. 303 pp. Price, bound: 13 roubles, 75 kopecks.

Struve, W. J., Studies on Stellar Astronomy. Translated by Prof. M. S. Eigenson. Ed. Corr. Member Acad. Sci. USSR A. A. Mikhailov (Series "Classics of Science") 1953. 234 pp. one appendix Price, bound: 10 roubles, 50 kopecks.

Reports of the Third Conference on Cosmogonic Problems, May 14-15, 1953.

The Origin of Cosmic Rays, 1954. N. A. Kisliakova. 1954. 230 pp., ill. Price, bound: 12 roubles, 75 kopecks.

Moscow, Kuibyshev Street, 8; Office "Akademkniga."

SOVIET ASTRONOMY — AJ VOL. 1, NO. 1 JANUARY-FEBRUARY, 1957

(Translation of the Astronomical Journal of the USSR Vol. 34, No. 1, published by the American Institute of Physics, September, 1958)

	Page	Russ. page
The Inclination to the Galactic Equator of the General Magnetic Field of the Galaxy in the Solar Vicinity. G. A. Shain	1	3
The Morphology of Galaxies. III. Diffuse Matter in Spherical Stellar Systems. B. A. Vorontsov-Vel'iaminov	9	8
The Determination of the Absolute Magnitudes of Cepheids and Supergiants and the Zero Point of the Period-Luminosity Relation. Iu. P. Pskovskii	19	19
The Distribution of Absorbing Matter and the Spiral Structure of the Galaxy. N. E. Kurochkin	31	31
On the Distribution of Mass in the Galaxy. A. M. Mikisha and F. A. Tsitsin	45	45
Permanent Configurations in the Problem of Four Bodies and their Stability. V. A. Brumberg	57	55
On Some Peculiarities in the Motions of Comets. P. G. Dukhnovskii	80	75
The Distribution of Comets of Jupiter's Family. K. A. Shteins	92	86
The Collimation of a Transit Instrument. G. P. Pil'nik	103	97
Increase of the Sensitivity of Some Photographic Emulsions by Means of Preliminary Short Exposure. T. M. Muliarchik and K. I. Petrova	108	102
The Early Thermal History of the Earth. V. G. Fesenkov	112	105
The Physical Nature of the Differences Between Geomagnetic Disturbances of Sudden and Nonsudden Commencement. E. R. Mustel'	124	120
A Possible New Type of Fluorescence of the Earth's Atmosphere. I. S. Shklovskii	130	127
Observations of Hydrogen Emission in Aurorae. Iu. I. Gal'perin	133	131

Brief Communications

	Page	Russ. page
On the Utilization of a Thin Lummer-Gehrke Plate. L. A. Fedin	138	135
A Numerical Method for Determining the Stellar Density in a Spheroidal Cluster. A. S. Sharov	142	139

Chronicle

	Page	Russ. page
Seventh Conference on Meteorites	144	141

SIGNIFICANCE OF ABBREVIATIONS MOST FREQUENTLY ENCOUNTERED IN SOVIET ASTRONOMY PERIODICALS

AN SSSR	*Academy of Sciences, USSR*
FIAN	*Physics Institute, Academy of Sciences USSR*
GAFI	*State Astrophysical Institute*
GAISh	*P. K. Shternberg State Institute of Astronomy*
GAO	*Main Astronomical Observatory*
Gidrometizdat	*Hydrometeorological Press*
GITI	*State Scientific and Technical Press*
GITTL	*State Press for Technical and Theoretical Literature*
GOI	*State Optical Institute*
GONTI	*State United Scientific and Technical Press*
Gosfizkhimizdat	*State Physical Chemistry Press*
GOST	*All-Union State Standard*
Goztekhizdat	*State Technical Press*
GTTI	*State Technical and Theoretical Press*
IF KhI	*Institute of Physical Chemistry Research*
IFP	*Institute of Physical Problems*
IL	*Foreign Literature Press*
IPF	*Institute of Applied Physics*
IPM	*Institute of Applied Mathematics*
ISN (Izd. Sov. Nauk)	*Soviet Science Press*
IIaP	*Institute of Nuclear Studies*
Izd	*Press (publishing house)*
KISO	*Solar Research Commission*
LFTI	*Leningrad Institute of Physics and Technology*
LIM	*Leningrad Institute of Metals*
LITMiO	*Leningrad Institute of Precision Instruments and Optics*
Mashgiz	*State Scientific-Technical Press for Machine Construction Literature*
MATI	*Moscow Aviation Technology Institute*
MGU	*Moscow State University*
Metallurgizdat	*Metallurgy Press*
MOPI	*Moscow Regional Pedagogical Institute*
NIAFIZ	*Scientific Research Association for Physics*
NIFI	*Scientific Research Institute of Physics*
NIIMM	*Scientific Research Institute of Mathematics and Mechanics*
NII ZVUKSZAPIOI	*Scientific Research Institute of Sound Recording*
NIKFI	*Scientific Institute of Motion Picture Photography*
OIIaI	*Joint Institute of Nuclear Studies*
ONTI	*United Scientific and Technical Press*
OTI	*Division of Technical Information*
OTN	*Division of Technical Science*
RIAN	*Radium Institute, Academy of Sciences of the USSR*
Ts NIIGAiK	*Central Scientific Research Institute of Geodetics, Aerial Survey, and Cartography*
URALFTI	*Ural Institute of Physics and Technology*
VAGO	*All-Union Astronomic and Geodetic Society*
VNIIM	*All-Union Scientific Research Institute of Meteorology*

NOTE: Abbreviations not on this list and not explained in the translation have been transliterated, no further information about their significance being available to us.—*Publisher.*

The expanded program of the American Institute of Physics comprises translation of six leading Soviet physics journals, as listed below. These translations, by competent, qualified scientists, provide all research laboratories and libraries with accurate and up-to-date information of the results of research in the U.S.S.R.

Soviet Physics—Technical Physics

A translation of the "Journal of Technical Physics" of the Academy of Sciences of the U.S.S.R. The translation began with the 1956 issues. Twelve issues per year, approximately 3,000 Russian pages. Annually $75.00 domestic, $79.00 foreign. Libraries* $35.00 domestic, $39.00 foreign. Back numbers, all issues $8.00.

Soviet Physics—Acoustics

A translation of the "Journal of Acoustics" of the Academy of Sciences of the U.S.S.R. The translation began with the 1955 issues. Four issues per year, approximately 400 Russian pages. Annually $12.00 domestic, $14.00 foreign. (No library discounts.) Back numbers, all issues $4.00.

Soviet Physics—Doklady

A translation of all of the "Physics Sections" of the Proceedings of the Academy of Sciences of the U.S.S.R. The translation began with the 1956 issues. Six issues per year, approximately 1,500 Russian pages. Annually $35.00 domestic, $38.00 foreign. Libraries* $15.00 domestic, $18.00 foreign. Back numbers for Volumes 1 and 2, $5.00 per issue; Volume 3 and subsequent, $7.00 per issue.

Soviet Physics—JETP

A translation of the "Journal of Experimental and Theoretical Physics" of the Academy of Sciences of the U.S.S.R. The translation began with the 1955 issues. Twelve issues per year, approximately 4,000 Russian pages. Annually $75.00 domestic, $79.00 foreign. Libraries* $35.00 domestic, $39.00 foreign. Back numbers, all issues $8.00.

Soviet Physics—Crystallography

A translation of the journal "Crystallography" of the Academy of Sciences of the U.S.S.R. The translation began with the 1957 issues. Six issues per year, approximately 1,000 Russian pages. Annually $25.00 domestic, $27.00 foreign. Libraries* $10.00 domestic, $12.00 foreign. Back numbers, all issues $5.00.

Soviet Astronomy—AJ

A translation of the "Astronomical Journal" of the Academy of Sciences of the U.S.S.R. The translation began with the 1957 issues. Six issues per year, approximately 1,100 Russian pages. Annually $25.00 domestic, $27.00 foreign. Libraries* $10.00 domestic, $12.00 foreign. Back numbers, all issues $5.00.

Subscriptions should be addressed to the American Institute of Physics, 335 East 45th Street, New York 17, N. Y.

---

* For libraries of non-profit academic institutions.